

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Facultade de Ciencias da Educación

Departamento de Métodos de Investigación e Diagnóstico
en Educación

ELABORACIÓN DUN TEST PARA ESTIMA-LO TAMAÑO DO
VOCABULARIO COÑECIDO EN LINGUA GALEGA

TESE DE DOUTORAMENTO

Autor: Antonio Estévez Álvarez

Director: Dr. Andrés Suárez Yáñez

Santiago de Compostela, 2007

Este traballo foi realizado na súa fase final durante unha licenza por estudos retribuída pola Consellería de Educación e Ordenación Universitaria

Agradecementos

Ó director da tese, Andrés Suárez Yáñez, porque soubo axudarme a abrir un espazo na investigación e por dirixi-la presente tese con rigor. Pero tamén pola constante preocupación que amosou e polo apoio continuado que me proporcionou. Non podo esquecer tampouco a manifestación de amizade que supón a atención que me adicou.

A María porque soubo prestarme o apoio que precisei en cada momento.

Ó profesor Xosé Rubal polas súas valiosas suxestións en diferentes ocasións.

Ás inspectoras de educación D.^a M.^a Asunción Fernández-Albor e D.^a Milagros Couto que me facilitaron o acceso ós centros de titularidade privada.

Ós profesionais da orientación e ás mestras que aceptaron colaborar na administración do test, así como ós docentes que participaron en calidade de xuíces e que non é posible nomear individualmente, pola súa contribución na realización da investigación.

Ó alumnado que participou porque a súa cooperación fixo posible que se puideran recaba-los datos necesarios para levar a cabo a investigación.

Índice xeral

Índice de táboas	xi
Índice de figuras	xvii
Introdución xeral	1

PRIMEIRA PARTE

Cuestións contextualizadoras

Introdución	7
-------------------	---

Capítulo 1. Delimitacións conceptuais e relevancia

1.1. Acerca do status lingüístico da <i>palabra</i>	13
1.1.1. Unha situación paradoxal	13
1.1.2. Léxico e vocabulario. Palabra e vocábulo	25
1.1.3. Palabra, oración, texto e situación	27
1.2. ¿Que significa coñecer unha palabra?	31
1.2.1. Unha primeira aproximación	31
1.2.2. Vocabulario activo e vocabulario pasivo	44
1.2.3. Vocabulario frecuente	47
1.2.4. Vocabulario dispoñible	61
1.2.5. Vocabulario usual	62
1.2.6. Vocabulario específico	64
1.2.7. Vocabulario oral e escrito	64
1.3. Relevancia do vocabulario	67
1.3.1. Correlación coa intelixencia medida polos tests	70
1.3.2. Lingua oral	72
1.3.3. Aprendizaxe inicial da lingua escrita	74
1.3.4. Comprensión lectora	76
1.3.5. Composición escrita	83

Capítulo 2. Estudos sobre o tamaño do vocabulario

2.1. Adquisición das primeiras palabras e desenvolvemento do vocabulario	91
2.1.1. Mecanismos e principios de adquisición de palabras	92
2.1.2. Organización e almacenamento das palabras	99
2.1.3. Ritmo de adquisición e evolución cuantitativa e cualitativa	102
2.1.4. Instrución en vocabulario	107
2.2. Tipos de estudos sobre tamaño do vocabulario	110
2.2.1. Recontos do vocabulario activo absoluto	111
2.2.2. Estimacións do vocabulario absoluto coñecido	114
2.2.3. Estimacións do vocabulario activo coñecido	115
2.3. Breve descrición de estudos de vocabulario en castelán e galego	115
2.3.1. García Hoz (1946)	116
2.3.2. Justicia (1985a, 1985b, 1995)	118
2.3.3. Casanova e Rivera (1989)	122
2.3.4. Esteban (1997)	125
2.3.5. Suárez, Seisdedos e Meara (1998)	127

Capítulo 3. Tests de vocabulario e a súa utilidade

Introdución: enfoques e procedementos na avaliación do vocabulario	131
3.1. Tests de tamaño relativo	134
3.1.1. Tests de intelixencia xeral e aptitudes	134
3.1.2. Tests específicos de vocabulario	137
3.2. Tests de estimación do tamaño absoluto	141
3.2.1. García Hoz (1946, 1976)	141
3.2.2. Test de Diack (1975)	141
3.2.3. Tests de resposta Si/Non	143
3.3. Medidas cualitativas ou de profundidade	160
3.3.1. Escala de Dale (1965)	161
3.3.2. Proposta de Read (1987)	162
3.3.3. <i>Vocabulary Knowledge Scale</i> (Wesche e Paribakht, 1993, 1996)	163
3.3.4. <i>Associates Vocabulary Test</i> (Read, 1993)	167
3.3.5. <i>Three Words Association Vocabulary Test</i> (Vives Boix, 1995)	170

SEGUNDA PARTE
Construción dun test de tamaño
do vocabulario en galego

Capítulo 4. Primeiras fases do proceso de construción do test

Introdución	179
4.1. A problemática acerca da construción de tests de tipo Si/Non	181
4.1.1. Decidir que contar como unha palabra	181
4.1.2. Elección das palabras para o test	183
4.1.3. Decidir como efectua-la avaliación do coñecemento do vocabulario	187
4.2. Elección dun dicionario	189
4.2.1. Características xerais dos dicionarios	189
4.2.2. Elección do dicionario da Real Academia Galega (RAG)	190
4.3. Selección das palabras e xeración de pseudopalabras do test	196
4.4. Sobre a corrección e puntuación	200

Capítulo 5. Administración do test e xustificación estatística

5.1. Aplicación piloto	207
5.2. Selección da mostra. Situación lingüística en Galicia	210
5.3. Administración definitiva. Mostra resultante	234
5.4. Fiabilidade	241
5.5. Validez	248
5.5.1. Consideracións sobre a validez	248
5.5.2. Validez de construto	252
5.5.3. Outros tipos de validez	280

Capítulo 6. Elaboración de baremos. Suxestións para a súa utilización

6.1. Datos descritivos básicos	288
6.2. Baremos	295

Capítulo 7. Conclusións	301
--------------------------------------	------------

Bibliografía	309
---------------------------	------------

Anexos

I. Instrucións ofrecidas ós xuíces para valora-la dificultade de aprendizaxe das palabras	337
II. Táboas coas pseudopalabras incluídas nos 10 subtests e o número de orde que ocupan en cada un	341
III. Formato inicial do test: Batería 1 e Batería 2	347
IV. Táboa para calcula-la estimación do tamaño do vocabulario	355
V. Instrucións iniciais para a aplicación do test	359
VI. Instrucións definitivas para a aplicación do test T-VOGAL	367
VII. Formato definitivo do test <i>T-VOGAL</i>	373
VIII. Táboas coas porcentaxes de sinalamentos dos ítems	383
IX. Táboas de baremos en centís	397

Índice de táboas

Táboa 2.1.	Porcentaxe de vocábulos en diferentes categorías, segundo Justicia (1985a)	105
Táboa 2.2.	Desenvolvemento do vocabulario en número de palabras, segundo Smith (1926)	113
Táboa 2.3.	Porcentaxe de palabras que coñecen os escolares (García Hoz, 1946)	117
Táboa 2.4.	Número de vocábulos que constitúen o vocabulario usual por idades, segundo Justicia (1985a, 1985b)	120
Táboa 2.5.	Número de vocábulos que constitúen o vocabulario usual por sexo, segundo Justicia (1985a, 1985b)	120
Táboa 2.6.	Media de palabras ou <i>tokens</i> utilizados en función do sexo, segundo Justicia (1995)	121
Táboa 2.7.	Número de vocábulos que constitúen o vocabulario usual por clase social, segundo Justicia (1985a, 1985b)	121
Táboa 2.8.	Media de palabras ou <i>tokens</i> utilizados en función da clase social, segundo Justicia (1995)	121
Táboa 2.9.	Número de vocábulos que constitúen o vocabulario usual por zona xeográfica, segundo Justicia (1985a, 1985b)	121
Táboa 2.10.	Puntuacións medias brutas e ponderadas do vocabulario pasivo coñecido (Casanova e Rivera, 1989)	124
Táboa 2.11.	Puntuacións medias brutas e ponderadas do vocabulario pasivo coñecido por escolares galegos (Esteban, 1997)	127
Táboa 2.12.	Estimación do número de palabras coñecidas, segundo EVOCA	129
Táboa 3.1.	Porcentaxe de palabras coñecidas a partir de Acertos e Erros (Anderson e Freebody, 1983)	154
Táboa 4.1.	Intervalos de confianza para diferentes tamaños de mostrás	186
Táboa 4.2.	Número de palabras incluídas en cada subtest	198
Táboa 4.3.	Exemplificación de datos utilizados para a estimación do vocabulario	202
Táboa 5.1.	Estimación do termo medio de palabras coñecidas. Primeira aplicación piloto	208

Táboa 5.2. Estimación do termo medio de palabras coñecidas por escolares de 6.º de Educación Primaria. Segunda aplicación piloto	209
Táboa 5.3. Utilización da lingua galega entre a poboación de 16 a 25 anos (rural/urbano)	213
Táboa 5.4. Dominio da lingua galega da poboación entre 16 e 25 anos (rural/urbano)	214
Táboa 5.5. Dominio da lingua galega da poboación xeral (rural/urbano)	215
Táboa 5.6. Dominio da lingua galega polo alumnado de 8.º de EXB (rural/urbano)	216
Táboa 5.7. Utilización da lingua galega entre a poboación de 16 a 25 anos (costa/interior)	218
Táboa 5.8. Utilización da lingua galega entre a poboación de 16 a 25 anos (nivel socioeconómico)	220
Táboa 5.9. Dominio da lingua galega da poboación xeral (nivel socioeconómico)	222
Táboa 5.10. Dominio da lingua galega do alumnado de 8.º de EXB (nivel socioeconómico)	223
Táboa 5.11. Dominio da lingua galega da poboación xeral segundo a lingua inicial	224
Táboa 5.12. Dominio da lectura en galego segundo a lingua inicial na poboación entre 16 e 25 anos	225
Táboa 5.13. Dominio da lingua galega segundo a lingua habitual na poboación xeral	226
Táboa 5.14. Poboación e mostra de 6.º de Educación Primaria	231
Táboa 5.15. Poboación e mostra de 2.º de ESO	231
Táboa 5.16. Resumo das mostras calculadas para 6.º de Educación Primaria e 2.º de ESO	232
Táboa 5.17. Número de centros que compoñen a mostra total	235
Táboa 5.18. Número de escolares da mostra total inicial	236
Táboa 5.19. Escolares excluídos e incluídos na mostra final de 6.º de Educación Primaria	238
Táboa 5.20. Escolares excluídos e incluídos na mostra final de 2.º de ESO	238
Táboa 5.21. Número de escolares que compoñen a mostra total definitiva	239
Táboa 5.22. Centros de procedencia do alumnado de 6.º de Educación Primaria segundo o hábitat e a titularidade do centro	240

Táboa 5.23. Centros de procedencia do alumnado de 2.º de ESO segundo o hábitat e a titularidade do centro	240
Táboa 5.24. Correlacións utilizando a estimación do número de palabras coñecidas para a mostra total	245
Táboa 5.25. Correlacións utilizando a estimación do número de palabras coñecidas para a mostra de 6.º de Educación Primaria	246
Táboa 5.26. Correlacións utilizando a estimación do número de palabras coñecidas para a mostra de 2.º de ESO	247
Táboa 5.27. Estatísticos descritivos da estimación do vocabulario coñecido para a mostra total	256
Táboa 5.28. Proba de Kolgomorov-Smirnov para o Total do Test na mostra total	256
Táboa 5.29. Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 6.º de Educación Primaria	257
Táboa 5.30. Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 2.º de ESO	258
Táboa 5.31. Contraste de Levene sobre homoxeneidade das varianzas (nivel de escolarización e hábitat rural vs urbano)	259
Táboa 5.32. Contraste de Levene sobre homoxeneidade das varianzas (nivel de escolarización e hábitat costa vs interior)	259
Táboa 5.33. Contraste de Levene sobre homoxeneidade das varianzas (nivel de escolarización e nivel socioeconómico-titularidade do centro)	259
Táboa 5.34. Contraste de Levene sobre homoxeneidade das varianzas (nivel de escolarización e lingua familiar habitual)	260
Táboa 5.35. Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Bat. 1)	262
Táboa 5.36. Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Bat. 2)	263
Táboa 5.37. Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Test Total)	264
Táboa 5.38. Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Bat. 1)	266
Táboa 5.39. Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Bat. 2)	267

Táboa 5.40. Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Test Total)	268
Táboa 5.41. Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Bat. 1)	270
Táboa 5.42. Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Bat. 2)	271
Táboa 5.43. Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Test Total)	272
Táboa 5.44. Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Bat. 1)	274
Táboa 5.45. Comparación post hoc (Scheffé) en nivel de escolarización e lingua habitual (Bat. 1)	274
Táboa 5.46. Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Bat. 2)	275
Táboa 5.47. Comparación post hoc (Scheffé) en nivel de escolarización e lingua habitual (Bat. 2)	276
Táboa 5.48. Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Test Total)	277
Táboa 5.49. Comparación post hoc (Scheffé) en nivel de escolarización e lingua habitual (Test Total)	277
Táboa 5.50. Resume da significación das diferenzas entre os distintos agrupamentos	280
Táboa 5.51. Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra total (Bat. 1, 2 e Test Total)	283
Táboa 5.52. Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra total (subtests)	283
Táboa 5.53. Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 6.º de Educación Primaria (Bat. 1, 2 e Test Total)	284
Táboa 5.54. Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 6.º de E. Primaria (subtests)	284
Táboa 5.55. Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 2.º de ESO (Bat 1, 2 e Test Total)	284
Táboa 5.56. Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 2.º de ESO (subtests)	285

Táboa 6.1.	Estatísticos descritivos da estimación do vocabulario coñecido nas mostras de 6.º de Educación Primaria e 2.º de ESO	288
Táboa 6.2.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra total para os ámbitos rural e urbano	289
Táboa 6.3.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 6.º de Educación Primaria para os ámbitos rural e urbano	289
Táboa 6.4.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 2.º de ESO para os ámbitos rural e urbano	289
Táboa 6.5.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra total para os ámbitos costa e interior	290
Táboa 6.6.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 6.º de Educación Primaria para os ámbitos costa e interior	291
Táboa 6.7.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 2.º de ESO para os ámbitos costa e interior	291
Táboa 6.8.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra total segundo a titularidade do centro (nivel socioeconómico)	292
Táboa 6.9.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 6.º de Educación Primaria segundo a titularidade do centro (nivel socioeconómico)	292
Táboa 6.10.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 2.º de ESO segundo a titularidade do centro (nivel socioeconómico)	292
Táboa 6.11.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra total segundo a lingua familiar habitual (galego vs castelán)	293
Táboa 6.12.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra total segundo a lingua familiar habitual (galego vs as dúas linguas)	294
Táboa 6.13.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 6.º de Educación Primaria segundo a lingua familiar habitual (galego vs castelán)	294
Táboa 6.14.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 6.º de Educación Primaria segundo a lingua familiar habitual (galego vs as dúas linguas)	294
Táboa 6.15.	Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 2.º de ESO segundo a lingua familiar habitual (galego vs castelán)	295

Táboa 6.16. Estatísticos descritivos da estimación do vocabulario coñecido na mostra de 2.º de ESO segundo a lingua familiar habitual (galego vs as dúas linguas)	295
---	-----

Índice de figuras

Figura 1.	Organización alternativa do contido da tese	4
Figura 1.1.	Supostos do coñecemento dunha palabra (Richards, 1976)	35
Figura 1.2.	Supostos en que se sustenta a quinta afirmación de Richards (1976)	36
Figura 1.3.	Resumo dos aspectos acerca do coñecemento dunha palabra	36
Figura 1.4.	Niveis de coñecemento dunha palabra na <i>VKS</i> (Wesche e Paribakth, 1996) ..	37
Figura 1.5.	Comparación dos procesos de lectura e escritura, segundo Cooper (1997)	84
Figura 2.1.	Trazado especulativo do incremento do léxico baseado en datos de Smith (1926), segundo Aguado (1999)	104
Figura 3.1.	Palabras incluídas nos niveis 1 e 6 do test número 8 de Diack (1975)	142
Figura 3.2.	Escala de coñecemento dunha palabra (Zimmerman et al, 1977)	145
Figura 3.3.	Comparación entre as puntuación en <i>First Certificate</i> e nun test Si/Non	151
Figura 3.4.	Comparación entre as puntuacións en <i>First Certificate</i> e nun test de elección múltiple	152
Figura 3.5.	Etapas no coñecemento dunha palabra (Dale, 1965)	162
Figura 3.6.	Escala de valoración do coñecemento dunha palabra (Wesche e Paribakth, 1996)	164
Figura 3.7.	Niveis "reais" de coñecemento dunha palabra na Escala de Wesche e Paribakth (1996)	166
Figura 3.8.	Ítem número 50 de <i>Associates Vocabulary Test</i> (Read, 1993)	168
Figura 3.9.	Test AVT de Vives Boix (1995), utilizado no estudio piloto	171
Figura 3.10.	Test A3VT de Vives Boix (1995), definitivo	171
Figura 4.1.	Gráfico da estimación do vocabulario coñecido a partir de Acertos e Erros	203
Figura 5.1.	Utilización da lingua galega entre a poboación de 16 a 25 anos (rural/urbano)	213
Figura 5.2.	Dominio da lingua galega da poboación entre 16 e 25 anos (rural/urbano)	214
Figura 5.3.	Dominio da lingua galega da poboación xeral (rural/urbano)	215

Figura 5.4.	Dominio da lingua galega polo alumnado de 8.º de EXB (rural/urbano) ...	216
Figura 5.5.	Utilización da lingua galega entre a poboación de 16 a 25 anos (costa/interior)	218
Figura 5.6.	Utilización da lingua galega entre a poboación de 16 a 25 anos (nivel socioeconómico)	220
Figura 5.7.	Dominio da lingua galega da poboación xeral (nivel socioeconómico)	222
Figura 5.8.	Dominio da lingua galega do alumnado de 8.º de EXB (nivel socioeconómico)	223
Figura 5.9.	Dominio da lingua galega da poboación xeral segundo a lingua inicial	224
Figura 5.10.	Dominio da lectura en galego segundo a lingua inicial na poboación entre 16 e 25 anos	225
Figura 5.11.	Dominio da lingua galega segundo a lingua habitual na poboación xeral ...	226
Figura 5.12.	Distribución da estimación do vocabulario para o Total do Test na mostra total	257
Figura 5.13.	Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Bat. 1)	262
Figura 5.14.	Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Bat. 2)	263
Figura 5.15.	Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Test Total)	264
Figura 5.16.	Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Bat. 1)	266
Figura 5.17.	Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Bat. 2)	267
Figura 5.18.	Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Test Total)	268
Figura 5.19.	Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Bat. 1)	270
Figura 5.20.	Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Bat. 2)	271
Figura 5.21.	Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Test Total)	272

Figura 5.22. Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Bat. 1)	275
Figura 5.23. Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Bat. 2)	276
Figura 5.24. Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Test Total)	278

Introdución xeral

A linguaxe verbal é un campo que sempre me interesou ó longo da miña traxectoria profesional como docente, tanto no ensino ordinario coma no ámbito das necesidades especiais, o que me levou a afondar no coñecemento acerca dos mecanismos e procesos que rexen a súa adquisición e desenvolvemento, así como nas dificultades e trastornos que presenta. Posteriormente, desempeñando a función de asesor de formación de profesorado en exercicio, fundamentalmente do especializado en n.e.e., tiven a oportunidade de constata-la importancia que lle conceden os docentes á linguaxe verbal, sendo a súa principal demanda a formación en contidos relacionados coa intervención en dificultades, tanto de lingua oral coma escrita. Isto constituíu un novo incentivo para seguir afondando neste campo, á vez que me permitiu establecer contacto con profesionais que me proporcionaron a oportunidade de amplia-la bagaxe de coñecementos. Nos últimos anos, exercendo como orientador no ensino primario, podo comprobar de xeito cotián que un dos aspectos nos que os docentes formulan peticións de axuda en maior medida resulta se-lo da linguaxe verbal en termos xenéricos, o que tamén se converte nun acicate, en canto que me esixe unha constante busca de información, para poder, cando menos, intentar responder ás demandas.

Ó mesmo tempo tempo, como persoa que ten o galego como lingua materna, preocúpame a situación do alumnado con esta lingua inicial. Por un lado, en termos xerais, temos que os escolares galegofalantes obteñen un rendemento lingüístico menor que os castelanfalantes (Brandín, 2002; Gómez, 1988; González Lorenzo, 1981; Pérez Pardo, 1985), se ben este dato está condicionado polo feito de que o alumnado galegofalante habitualmente tamén é o que pertence a clases sociais menos favorecidas social e

economicamente, como xa apuntou Pérez Vilariño (1981) en relación co rendemento escolar. Pero, tamén é certo que este feito non sempre parece representar unha preocupación para o colectivo docente, como o é que o coñecemento da lingua galega segue a considerarse como complementario, pero non substancial, aínda que a situación sexa máis vantaxosa que a que mostraban os datos fai máis de vinte anos (Arza, Campos, Veiga e Rubal, 1990; Fernández, 1983; Pérez Vilariño, 1979; Ramírez, 1973; Rojo, 1981) ou máis de quince (Arza, Campos, Veiga e Rubal, 1990; Cajide, 1989; Rubal, 1988). Non obstante, ó meu xuízo, si se pode entender que a consideración da lingua galega no entorno escolar, como probablemente se podería dicir da súa consideración social, continúa sen ser precisamente alta.

O profesorado, cando menos nos niveis de ensino obrigatorio, adoitamos te-la convicción, aínda que habitualmente só sexa unha intuición, de que, dentro da linguaxe verbal, o vocabulario é un dos piares en que se asenta a aprendizaxe do alumnado; que cando o seu tamaño é limitado, pobre, seguramente se presentarán dificultades para aprender, polo que é necesario incrementa-lo seu volume (outra cuestión son as concepcións acerca de como se aprende ou como se aborda a súa ensinanza). Existe a impresión de que a maior tamaño do vocabulario, maior é o volume de coñecementos, maiores son as posibilidades de comprensión dos contidos académicos, así como a de textos. Esta impresión está en consonancia coa seguinte aseveración de Kintsch (1998, p. 165): "A construción dos significados dun texto non remata coa construción dos significado das palabras, pero empeza nela". Se é así, parece razoable pensar que pode ser útil valorar cal é o grao de coñecemento de que dispoñen os escolares. Existen probas en español que posibilitan esta valoración, pero non sucede o mesmo en galego para persoas galegas, sexan galegofalantes ou non.

En diferentes momentos do meu periplo profesional, utilicei distintas probas de vocabulario en castelán tanto con alumnado de ensino ordinario coma con alumnado con necesidades especiais. Entre elas, algunhas teñen por finalidade situa-lo suxeito nun baremo

de idade, sexan exclusivamente de vocabulario, aínda que tamén posibiliten obter información de que palabras coñecen ou non (a de García Hoz de 1976 e o Test de Vocabulario en Imáxenes de Peabody), ou incluídas en tests de aptitude intelectual con diferentes características (TEA, WISC-R e BAPAE), así como outras non estritamente de vocabulario, pero moi relacionadas, como é o Test BOEHM de Conceptos Básicos. Na etapa dos cursos de doutoramento, tiveren acceso a outro tipo de tests, como é o EVOCA (Suárez, Seisdedos e Meara, 1998), que pretende estima-lo volume de vocabulario que coñece unha determinada persoa, e require respostas de tipo *Sí/Non coñezo a palabra que se presenta*; posteriormente, tiveren a oportunidade de profundar no seu coñecemento e xurdiu a idea de construír unha proba destas características en lingua galega e para alumnado galego. A idea foi madurando, ata adoptar-la decisión de plasmala na práctica, e, como consecuencia, se desenvolveu este traballo.

O propósito xeral do mesmo, polo tanto, é construír unha proba a partir da información achegada en traballos como os de García Hoz (1946), Meara e Buxton (1987), Nation (1993a) e Suárez, Seisdedos e Meara (1998), entre outros, que, mediante cálculos baseados na inferencia, permita estimar de forma fiable e válida o vocabulario galego coñecido por escolares galegos de 8 a 16 anos, e poida servir como instrumento potencialmente útil na práctica pedagóxica e para a investigación.

O informe da investigación está organizado en dúas partes: unha está adicada á abordaxe de cuestións que permiten contextualiza-la achega empírica da segunda parte, e unha segunda parte na que se expón o proceso de construción do test para a estimación do tamaño do vocabulario.

Na primeira parte, o capítulo primeiro adícase á delimitación de conceptos e termos tales como o de palabra (léxico e vocabulario, funcións e tipos de palabras), que significa coñecer unha palabra e á súa relevancia. No segundo capítulo pártese da adquisición e da evolución do vocabulario e analízanse os tipos de estudos máis habituais sobre vocabulario,

á vez que se describen algúns deles. Finalmente, no terceiro capítulo realízase unha análise dos diferentes tipos de tests que teñen como finalidade a valoración do vocabulario e que, basicamente, son os normativos, os de tamaño e os que intentan ofrecer unha medida máis fonda do seu coñecemento.

En canto á segunda parte, no capítulo cuarto descríbese o proceso de construción das dúas baterías de que consta o test, dende a elección dun dicionario, pasando polo procedemento de selección de palabras, ata a propia construción do test. No quinto capítulo, relativo á administración do test, explícase o proceso e tarefas realizadas para a selección de mostras, para a aplicación das probas, a corrección destas e as análises de fiabilidade e validez. No capítulo sexto expóñense datos descritivos, o proceso de elaboración de baremos e suxestións para a súa utilización. Por último, no capítulo sétimo trátase de ofrecer unha serie de reflexións e de extraer conclusións, en base á información e ós datos conseguidos.

Finalmente, é posible realizar unha lectura do contido da tese seguindo unha orde diferente, ofrecendo unha perspectiva distinta do devandito contido. Con elo, trátase de contribuír á clarificación da súa estrutura (Figura 1).

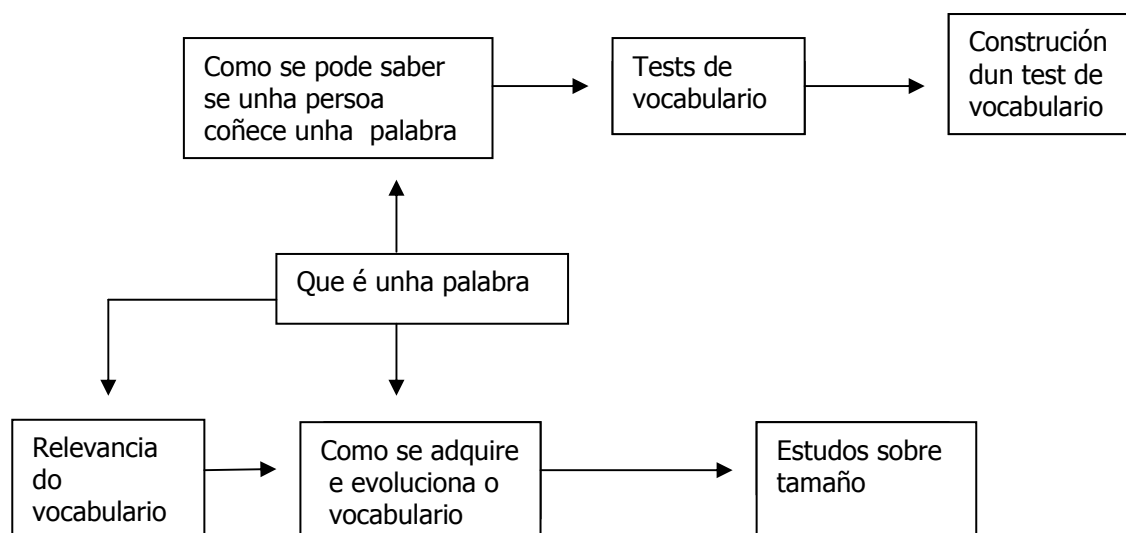


Figura 1.: Organización alternativa do contido da tese

PRIMEIRA PARTE

Cuestións contextualizadoras

Introdución

O vocabulario pode considerarse como a manifestación máis inmediata da linguaxe e do coñecemento do mundo que posúen os seres humanos, o cal constitúe a razón básica para concederlle importancia á investigación neste campo. Para autores como García Hoz (1946) ou Justicia (1985a) o interese polo seu estudo aséntase en argumentos de distinta índole. Un dos argumentos é de carácter científico, derivado da necesidade de alcanzar un mellor coñecemento en relación co ritmo en que se incrementa no transcurso do desenvolvemento evolutivo.

Outro argumento é de carácter práctico-pedagóxico. Pódese considerar que o vocabulario configúrase como un elemento da cultura, porque representa a expresión da riqueza mental, das ideas aprehendidas dun contido cultural e porque constitúe un vehículo de ensinanza/aprendizaxe, xa que coñece-lo significado dos termos utilizados é unha condición para a adquisición de coñecementos. Considerando o vocabulario coma un dos piares fundamentais nos que se apoia a ensinanza e a aprendizaxe escolar, coñece-lo ritmo de adquisición do mesmo e a competencia do alumnado ó respecto será un elemento que permitirá adecua-la intervención didáctica ás súas posibilidades en aspectos esenciais das diferentes áreas do currículo.

Un terceiro argumento no que fundamenta-lo valor da investigación neste aspecto da linguaxe ven dado polas correlacións obtidas coa intelixencia xeral en estudos como os de García Hoz (1946), Dunn (1985) ou no da adaptación española do WISC (1989, 1993), no que a subproba de vocabulario é das que mantén unha maior correlación co total da escala, pero tamén coa verbal. Por outra parte, isto parece algo aceptado na maioría dos tests de intelixencia, en tanto que adoitan concederlle un importante peso, incluíndo subprobas

relacionadas co vocabulario ou en probas de selección, como é o caso da *Cambridge First Certificate Examination* (Meara e Buxton, 1987), na que o vocabulario é un elemento fundamental.

Enténdese por vocabulario o conxunto de palabras que emprega e comprende efectivamente unha persoa ou unha comunidade lingüística, en contraposición a léxico que está constituído por tódalas palabras que compoñen unha lingua e están a disposición dos falantes da mesma, aínda que non as utilicen nin comprendan. En consecuencia, o tamaño do vocabulario dun individuo fai referencia ó número de vocábulos que ese individuo emprega e comprende en actos concretos de fala ou en textos escritos.

Segundo Nation (1993a), o estudo máis importante e antigo sobre tamaño de vocabulario é *The Vocabularies of School Pupils* de Thorndike, en 1924, no que revisou nove traballos nos que se empregaban dicionarios como fontes para a selección das palabras usadas para construí-las probas mediante as que se efectuaron as estimacións do vocabulario. Thorndike detectou unha ampla disparidade nas estimacións, que atribuíu a que os estudos non cumprían tres ou os catro requisitos que el consideraba imprescindibles e que se detallan deseguido. En primeiro lugar, utilizar un dicionario o suficientemente amplo que recolla o vocabulario que poidan coñecer-las persoas que participen na investigación, evitando, así, unha subestimación do vocabulario. En segundo lugar, delimitar que palabras serán susceptibles de ser seleccionadas, procurando non incluír palabras pertencentes a unha familia ou homógrafos, xa que ó aparecer nun maior número de entradas no dicionario, é máis probable que sexan seleccionadas e, con elo, prodúcese unha tendencia á sobreestimación do vocabulario. En terceiro lugar, o procedemento para selecciona-las palabras debe evitar que se produza unha tendencia a elixir-las máis frecuentes, que son as que ocupan máis espazo e das que teñen un maior número de entradas, como ocorre con procedementos consistentes en escoller a primeira palabra de cada páxina, porque tamén conducen a unha sobreestimación do vocabulario. En cuarto lugar, a mostra de palabras

debe conter un número adecuado de palabras dos diferentes niveis de frecuencia, con obxecto de evitar tamén con esta medida que as palabras de maior frecuencia teñan máis probabilidades de ser seleccionadas.

Posteriormente, Williams (1932) comparou unha mostra de palabras extraídas dun dicionario cun vocabulario infantil, consistente nun listado de vocablos que se consideraban frecuentes e básicos para unhas determinadas idades. Williams encontrou que na mostra de palabras do dicionario aparecían unha proporción de palabras da lista de vocabulario infantil moi superior á que se podería esperar, chegando a unha conclusión semellante á de Thorndike, no sentido de que o método de mostraxe de palabras e o tamaño do dicionario determinan a composición dunha mostra de palabras, de forma que, se o procedemento é inadecuado, as palabras de alta frecuencia obteñen unha maior representación da que lle correspondería.

En España pódese cita-lo traballo de García Hoz (1946) como o primeiro no que se trata de estima-lo tamaño do vocabulario con escolares españois. Xa nos anos setenta e oitenta apareceron estudos como os de Armayor (1975, 1977), Justicia (1983, 1985b, 1995) ou Lecuona (1988), entre outros.

Nas dúas últimas décadas produciuse tamén un incremento na preocupación polo vocabulario como un dos compoñentes básicos da competencia en ámbitos como a aprendizaxe dunha segunda lingua (Meara, 1987, 1993a, 1994; Meara e Buxton, 1987; Vives Boix, 1995, por exemplo) ou na expresión escrita (Carter, 1986; Giammateo e Basualdo, 2003, entre outros), cando tradicionalmente se acantoara en pro doutros aspectos da aprendizaxe da lingua, tales como a fonoloxía, morfoloxía ou sintaxe. A amplitude do seu tamaño e a aparente falta de estruturación relegouno a un lugar secundario no ensino da lingua e na investigación, de forma que a miúdo ten sido tratado como un apéndice máis que como unha parte esencial nos programas de ensino, agás os vocabularios básicos. O xurdimento dun maior interese favoreceu o desenvolvemento de enfoques de ensinanza e

aprendizaxe baseados no léxico en maior medida. Do mesmo xeito, produciuse un crecente interese en investigar cuestións básicas sobre a natureza do lexicón mental dos aprendices dunha segunda lingua, o desenvolvemento do mesmo e as semellanzas e diferenzas co dos falantes nativos.

Para darlle resposta a estas cuestións, fíxose preciso dispoñer de instrumentos fiables que permitiran avaliar tanto o desenvolvemento do léxico coma o coñecemento de palabras, xa que existen cuestións importantes por resolver sobre a cantidade e calidade do vocabulario que se necesita para aprender, como avalia-lo coñecemento do vocabulario antes e despois dun período de instrución ou o tamaño do vocabulario que coñecen os aprendices e en que grao de profundidade.

Os instrumentos de medida tradicionais, tales como os tests de elección múltiple, utilizan un limitado número de palabras exploradas con certa profundidade, pero non permiten ofrecer unha visión ampla do desenvolvemento do lexicón e a miúdo vese interferida por outras habilidades lingüísticas. Máis recentemente, téñense utilizado con efectividade técnicas máis simples para a estimación do tamaño de vocabulario total. Unha mostra son os tests de resposta Si/Non (Meara, 1987, 1990, por exemplo), amplamente experimentados en diferentes linguas. Outro interesante enfoque levárono a cabo autores como Read (1990), que propoñen unha medida da calidade do vocabulario, en base a tests de asociacións. Todos estes intentos verten luz sobre diferentes aspectos do coñecemento das palabras, pero é preciso ampliar e afondar no coñecemento específico sobre cal sería o procedemento óptimo para avalia-lo coñecemento de palabras.

A información relativa ó seu coñecemento resulta relevante na medida en que pode proporcionar datos que contribúan á comprensión de aspectos relacionados co éxito/fracaso escolar, detectar alumnado con vocabulario excesivamente pobre, planifica-la liña a partir da cal avaliar programas de intervención ou establece-lo ritmo de crecemento do vocabulario, por exemplo. Por esta razón, parece de interese dispoñer dunha referencia de utilidade,

tanto no ámbito do desenvolvemento do vocabulario como noutros relacionados con este, como son o da selección de materiais didácticos impresos ou da comunicación verbal na aula.

Con este propósito, como xa se sinalou na introdución xeral, é co que se planea neste traballo a elaboración dun instrumento mediante o cal sexa posible estima-lo vocabulario coñecido en lingua galega por escolares de Galicia.

Capítulo 1

Delimitacións conceptuais e relevancia

Neste capítulo tratarase de contextualizar e clarificar dous conceptos fundamentais para a investigación que se presenta na segunda parte desta tese. Por un parte, o concepto de palabra (que se entende por esta unidade lingüística e a súa distinción co concepto de vocábulo), así coma outras cuestións relacionadas (que se entende por coñecer unha palabra e a importancia tanto do texto coma da situación na significación da mesma). Por outra parte, abórdase o concepto de vocabulario (que se entende por vocabulario e a súa diferenza con léxico), xunto con cuestións como son os distintos tipos de vocabulario e a súa relevancia.

1.1 Acerca do status lingüístico da *palabra*

A palabra é unha unidade lingüística como tamén son unidades lingüísticas o monema, o morfema, o sintagma, a oración e o texto. Na actualidade concédeselle unha maior importancia á unidade texto, pero pódese dicir que a unidade palabra, con tódalas dificultades que existen para a súa delimitación, é unha unidade básica que reflicte a construción que as persoas levan a cabo da realidade.

1.1.1 Unha situación paradoxal

Palabra é un termo que se utiliza con moita frecuencia tanto no contexto da vida cotiá como no da investigación e académico, e de feito nesta tese trabállase sobre a base de que a información acerca do número de vocábulos que coñece un escolar é unha información útil. Pero, como se verá a continuación, non é fácil definir este término. Son ben

coñecidas as dificultades para lograr unha definición válida para tódalas linguas (Blanche-Benveniste, 1998, pp. 76 e sgtes.; Fruyt e Reichler-Béguelin, 1990; Reichler-Béguelin, 1992), do mesmo xeito que é ben coñecido o rexeitamento dos lingüistas do século XX á utilización dunha noción, que, ó sumo, parece ter un status intuitivo, preteórico (Malkiel, 1970; Martinet, 1966, p. 51; Palmer, 1983, p. 32; Pergnier, 1986, p.16). Non obstante, a idea que podería denominarse común de *qué é unha palabra* parece simple, xa que todos sabemos que é e falamos poñendo palabras xuntas. É unha noción incerta, preteórica, intuitiva, pero seguramente insubstituíble, cunha existencia tanxible na escrita, onde na actualidade se separan as palabras por espazos en branco. Aínda que os lingüistas discutan a súa propia existencia, cando escriben ou falan sobre a palabra, fano usando e separando palabras. Martinet (1972) exprésao do seguinte modo:

Non obstante, é innegable que no plano psicolóxico é un elemento real e válido; o suxeito falante, aínda inculto e, segundo Sapir, calquera que sexa a estrutura da súa lingua, pensa e pronuncia palabras, de forma que os propios lingüistas vense obrigados a utiliza-la palabra no seu sentido tradicional e habitual (305).

A utilización habitual de *palabra* (en dicionarios, na vida cotiá, etc.) contrasta coa dificultade para definila con precisión como se comentou máis arriba. Para ilustrar esta situación, é de sinalar que se define frecuentemente a palabra facendo referencia á súa separación en espazos en branco na escritura, como ocorre con Justicia (1993, p. 129), quen, seguindo a Hocket (1958), a considera como “toda unidade gráfica ou conxunto de signos separado dos grupos veciños por espazos en branco ou algún signo de puntuación”. Pero, non sempre se separaron as palabras na escritura mediante espazos como se fai na actualidade. Este non se produciu de forma sistematizada ata o século VIII d.C. Anteriormente existía unha preferencia, que non inexistencia da escritura separada, pola

denominada *scriptio continua* ou escritura sen segmentación, que obrigaba ó lector a separar e puntuar o texto. Poderíase dicir que a lectura era practicamente unha interpretación do texto por parte do lector, como sucede coa interpretación dunha partitura musical. Non deben sorprender, polo tanto, as dificultades e oscilacións de calquera individuo para segmentar convencionalmente as palabras, ben sexa un neno, como coñece calquera docente que traballe con escolares que están iniciándose ou presenten dificultades na escritura, ben sexa un adulto que non domine o sistema de escritura (Suárez, 2000, p. 22).

Na linguaxe oral, a palabra tampouco é unha unidade lingüística básica. Así, desde o punto de vista da entoación, cando se escoita un discurso, un texto falado, máis que palabras percíbense *grupos fónicos*, que se definen como porcións do discurso comprendidas entre dúas pausas. Os grupos fónicos habitualmente oscilan entre 8 e 11 sílabas (*que-teña-boa-viaxe, hoxe-en-día*), pero hainos tamén dunha sílaba, como sucede, por exemplo, en respostas de *si* ou *non* a unha pregunta. Desde o punto de vista da fonosintaxe, que estuda as modificacións que sofren os fonemas ó agruparse coas palabras, dentro do período, fálase de *sirremas* (Quilis e Fernández, 1979, p. 143), agrupacións de dúas ou máis palabras que se producen pola necesidade de apoio acentual. Os sirremas constitúen unha unidade gramatical perfecta, unha unidade tonal, xa que os seus elementos non se separan por pausas entre eles, unha unidade de sentido e, ademais, forman a unidade sintáctica intermedia entre a palabra e a frase (*oínovento, canblanco, ocandeluis, deixamosdeser nós mesmos, pasaronben os máis destacados alumnos, Rosa e Carmen, a cama de Antonio*). Tamén se fala de *concorrenza de fonemas homólogos* (vogais e consoantes), que se produce cando ó final dunha palabra aparece un fonema e ó inicio da seguinte palabra repítese o mesmo fonema. Un exemplo de concorrenza é o fenómeno da *sinalefa*, o enlace de vogais homólogas ou diferentes (*a-abundancia, a-escola, veño-a-empezar, pouco-efecto*).

Desde unha perspectiva semántica, hai palabras como os nomes de clase natural (*león, mesa, rocha ...*) nos que é fácil dicir cal é o seu referente, o significado, pero non

sucede o mesmo no caso doutro tipo de palabras (palabras gramaticais, abstractas, frases feitas, etc.).

En cuarto lugar, que a palabra non é unha unidade lingüística claramente definida queda mostrado en casos como *sacarrollas*, en que non está claro se é unha ou son dúas as palabras, ou no caso de frases idiomáticas. Ademais, é posible que un mesmo significado se poda expresar cunha ou varias palabras (*valdeorrés=nado en Valdeorras; levar a cabo=realizar*); tamén ocorre que o que se di nunha lingua cunha palabra, noutra lingua utilízanse máis dunha palabra (*plancha=ferro de pasar; mazá=pomme de terre*).

Todos estes exemplos son indicadores de que non parece moi acertada a idea de que a palabra é unha unidade semántica natural, aínda que se utilice para a elaboración dos dicionarios. Os devanditos exemplos levan a pensar que o máis característico da palabra é a forma e non o significado, o que está en consonancia coa definición de palabra que da Bloomfield (1933) como “mínima **forma** libre”.

Saussure conservou o termo de “palabra”. Na súa teoría do signo, a palabra enténdese como un signo lingüístico, no que se diferenzan dous aspectos: significante e significado. Compara o signo cunha folla de papel con dúas caras que son inseparables aínda que se corte a folla en anacos. O significante é concibido como a imaxe acústica que se produce na mente cando se escoita unha palabra e que permite recoñecela cando se escoita de novo. O significante non é para Saussure a secuencia de sons que constitúen a palabra senón a imaxe acústica que producen na mente, diferenza semellante á que se pode establecer entre un libro real e a imaxe que dese libro se fai unha persoa cando pecha os ollos; con todo, no significante adóitase incluír tanto a sucesión de sons coma a imaxe acústica que producen. Un sinónimo de significante sería expresión, fronte a contido, que sería o significado.

Esta idea do signo lingüístico encádrase nun concepto de lingua como un sistema de signos. Tanto o signo coma a lingua son considerados en si mesmos, como entidades fixas e

descontextualizadas, resultado dunha convención social. Sublíñase como a función básica da linguaxe a función representativa (hai unha realidade que se copia mediante a linguaxe) e os individuos teñen que aprender os significantes e o seu significado. En relación con esta idea sitúase a convicción común da validez das palabras en si mesmas e a suposición da existencia de cousas detrás dos rótulos, así como a crenza implícita na realidade das ideas abstractas. Nalgúns casos, incluso, se recea do seu poder e aparecen tabús verbais (*morte, Deus, etc.*) que se deben evitar dicir. Non obstante, en canto se analiza detidamente, non está claro o status da palabra como signo lingüístico sustentado na idea a que se acaba de facer referencia: os signos fixados por convención que representan a realidade e que hai que aprender.

A concepción de palabra como signo lingüístico séguese a utilizar, pero é de resalta-la crítica da semioloxía integracionista (Harris, 2000) ó concepto de signo lingüístico como algo fixo e creado por convención social, subliñando que o importante, de onde se debe partir, é da capacidade das persoas para crear signos e significados (función simbólica), coa finalidade de resolve-las situacións que se lle presentan, cuestionando a existencia de signos descontextualizados, o que se pode considerar compatible co concepto de palabra de Bloomfield (mínima forma libre). As expresións/formas/significantes deben ser dotados de significación polo oínte ou lector en función da situación comunicativa. Pódese conectar igualmente con Vigotsky (1962, p. 5), para quen *word meaning* é a unidade que está presente tanto no pensamento como na linguaxe, representa conceptos, é o instrumento básico do pensamento e este vai sendo elaborado, vai sendo construído polo propio suxeito ó longo da vida. As palabras non son representantes das cousas. Son afirmacións que se refiren ás cousas e interpretámolas sen ser conscientes. Grijelmo (2000) exemplifícao como segue:

A palabra árbore correspóndese, pois, a unha infinidade de imaxes posibles que dependen na súa aplicación concreta do lugar onde se escoite o vocábulo, do clima, da natureza local, da súa bioloxía ... Porque a linguaxe humana non representa ou sinala directamente a realidade, senón que `representa representacións mentais que os suxeitos teñen e constrúen acerca da realidade´ (p. 116).

O término palabra é difícil de definir, se ben parece haber acordo en canto a que é unha unidade mínima ou case mínima. Pero, para algúns autores é unha unidade de pensamento, para outros unha unidade da lingua, para outros de fala e para outros de sons. Insístese con frecuencia na posibilidade de que se desprenda da cadea falada e funcione soa, que represente unha entrada do dicionario ou que constitúa unha palabra base, un enunciado por si soa.

A continuación expóñense algúns exemplos de intentos encamiñados a aclarar que é unha palabra e que reflicten a dificultade para definila. Un claro exemplo represéntao o dicionario da Real Academia Española que na edición de 1987 define as palabras como "son ou conxunto de sons articulados que expresan unha idea" (p. 996), mentres que na de 2001 dise que son o "segmento do discurso unificado habitualmente polo acento, o significado e pausas potenciais inicial e final" (p. 1650). Pola súa parte, o Diccionario da Real Academia Galega establece que as palabras "son ou conxunto de sons que expresan unha idea e constitúen unha unidade autónoma, así como a representación gráfica dos mesmos" (1997, p. 854).

Bloomfield (1933, p. 178) utilizou un criterio formal máis que semántico ó defini-la palabra nos términos sinalados máis arriba como "mínima forma libre". De modo semellante, para Alonso (1986, p. 68) é "a unidade mínima, independente e libre que ten significante e significado, composta por un ou máis monemas". Martinet (1966, p. 40), referíndose tanto ó plano oral como ó escrito, enténdea como "segmento da cadea falada ou dun texto escrito,

que se pode extraer do contexto, pronunciándoo illadamente ou separándoo por un espazo en branco dos demais elementos do texto e atribuírle un significado cunha función sintáctica". Para Vigotsky (1962, p. 4) a palabra é a unión de son e significado.

En relación coa conceptualización como mínima forma libre, a máis pequena forma que poida ocorrer illadamente, dependerá do que se pretenda dicir con "illadamente", xa que dificilmente se utilizarán palabras función/gramaticais de forma illada. Bloomfield suxire buscar un elemento máis pequeno que a palabra, unha unidade de significado que é o *morfema* (*monema* para os lingüistas europeos). Non obstante, en tanto que hai palabras que non é posible segmentar en partes con significado propio dun modo obvio (*houbo* componse de *haber* e o significado de *pasado*), para Palmer (1983, p. 34) é preferible redefini-lo término palabra de modo que, por exemplo, *ama* e *amado* son dúas formas da mesma palabra que é o verbo *amar*. O término para denominalo é *lexema*, que é o que habitualmente constitúen as entradas dos dicionarios. Carter (1986) entende que é máis adecuado traballar cos termos/conceptos de *ítems lexicais* ou *lexemas* que traballar co de palabra, xa que os primeiros permiten englobar frases lexicais, como sucede en "pasara a mellor vida", por morrer.

Martinet (1972) sinala que, mentres que *lexema* é un término que designa unha unidade de lexicografía que encontramos no léxico, o término palabra é impreciso en canto ó seu contido e pódelle interesar tanto a psicólogos como a lóxicos, a estruturalistas como a foneticistas ou a poetas (p. 302). Durante moito tempo considerouse como a unidade central de toda lingua, pero, tendo en conta a dificultade para definila, podería ser interesante substituí-lo termo no campo da investigación por "monema e sintagma, segundo os casos, xa que teñen a vantaxe de ser utilizados con referencia a tódalas estruturas lingüísticas" (p. 305).

Para Pinker (1994) esta confusión deriva de que o término palabra non é cientificamente preciso e refírese a dous conceptos. O primeiro fai referencia a *átomos*

sintácticos (p. 147), construídos a partir dos regulamentos da morfoloxía e que se comportan como as mínimas e indivisibles unidades sintácticas; inclúen formas de nomes, de verbos, derivados formados con sufixos, con prefixos e por composición, así como formas irregulares: coinciden co que Crystal (1995) denomina o primeiro tipo de lexemas, palabras derivadas morfoloxicamente. O segundo son os *listemas* (p. 142), unidades de significación que non se poden xerar mediante regulamentos morfolóxicos e é necesario memorizalas como ocorre no caso de locucións e frases feitas: correspóndense coa segunda categoría de lexemas de Crystal.

A teor do que se acaba de comentar ó intentar definir que é unha palabra, procede clarificar a que se fai referencia cando se alude a unha serie de unidades lingüísticas relacionadas coa palabra como son, segundo Alonso (1986):

Os *monemas* (*morfemas* en terminoloxía norteamericana), unidades mínimas con significación, indivisibles en canto ó significado en elementos menores (*árbore, tres, ben, Buenos Aires*, etc., pero tamén *en, que*, etc.), aínda que consten de máis dunha palabra, como é o caso de *Buenos Aires*, por exemplo. Tamén poden formar parte dunha palabra, de modo que algunhas palabras poden constar de máis dun monema, como sucede con *agrisado* (*a-*, dirección, tendencia; *gris*, un color; e *-o*, xénero masculino) ou *inútiles* (*in-*, negación; *-útil*, que serve para algo; e *-es*, plural).

Os *monemas*, á súa vez, divídense en *lexemas* e *morfemas*. Os *lexemas* (*morfemas léxicos* en terminoloxía norteamericana) son os monemas que teñen significado propio e que constitúen a raíz dos nomes, adxectivos, verbos e adverbios. Están nos dicionarios e o seu número é enorme en cada lingua. Por exemplo, en *falabas*, sería *fal-*, en *inutilizados*, sería *-útil*, en *visitaremos*, sería *visit-* e en *pescadores*, sería *pesc-*. Os *morfemas* (*morfemas gramaticais* en terminoloxía norteamericana) son os monemas que non teñen significado pleno, ou que o teñen moi diminuído; serven para expresa-las relacións entre lexemas, ou para completar dalgún modo a súa significación. Existen tres clases fundamentalmente:

- Independentes ou funcionais, que serven para representa-los lexemas ou para relacionalos entre si. Algúns exemplos son os determinantes, as preposicións, as conxuncións, os pronomes relativos, etc. Constitúen palabras soltas, aparecen nos dicionarios e estúdanse na gramática.

- Flexivos, desinencias ou afixos flexivos, que son elementos que serven para expresar xénero e número en nomes e adxectivos, así como modos, tempos, número e persoas dos verbos. As desinencias son a terminación variable da palabra e sinalan as concordancias dos nomes, adxectivos e verbos. Algúns exemplos son *-a-* e *-s* en *casas altas*, *-ba-* e *-mos* en *saltabamos*.

- Derivativos ou afixos derivativos, que son as partículas coas que se forman, xunto coa raíz, as palabras derivadas e compostas. A diferenza das desinencias, non son imprescindibles xunto ós nomes, adxectivos, verbos ou adverbios e non entran na concordancia. Exemplos son *-mente* en *fundamentalmente*, *re-* en *refacer*, *-os-* en *gracioso*, *-ísim-* en *grandísimo*, etc.

Outras dúas cuestións que merecen ser comentadas son, por un lado, a relativa á variedade de palabras existentes e, por outro, o concepto de familia de palabras en tanto que son relevantes no que se refire á elaboración de dicionarios e na selección de palabras que se utilicen para a construción de instrumentos de avaliación do tamaño do vocabulario, como o que se pretende elaborar nesta tese.

En canto ós **tipos de palabras** non hai acordo entre os autores. Os criterios utilizados para a súa clasificación son diversos e, en consecuencia, tamén o son as propias clasificacións. Alonso (1986), por exemplo, por un lado, clasifica as palabras segundo a forma en simples (monema=lexema+morfema) e compostas (dous ou máis lexemas), e, por outro, segundo a orixe, en primitivas e derivadas.

Carter (1986) establece unha primeira distinción entre ítems lexicais pechados e abertos, atendendo á súa función no texto. Os primeiros comprenden un rango moi limitado,

son aqueles que xogan un papel indispensable na cohesión dun texto, tales como son os artigos, conxuncións, pronomes e verbos auxiliares e ós que se lle denomina palabras gramaticais, baleiras, estruturais, función, etc. Entre os segundos, os abertos, están os substantivos, adxectivos, verbos e adverbios. O seu número é teoricamente infinito e tamén se lles denomina con termos como palabras lexicais ou palabras contido. O seu rol no texto é o de aportar información e son menos predicibles. Autores como Maratsos (1990) non aceptan esta clasificación e dividen as palabras en nomes (base referencial, aínda que sexa indirecta como a que posúen os pronomes) e as demais (base estrutural).

Unha segunda distinción que establece Carter (1986) baséase no tipo de significación que aportan as palabras, diferenciando entre nucleares (*core*) e non nucleares (*no core*). As nucleares ou inexpressivas son as que, en principio, unicamente se refiren ou denotan algo sen matices como ocorre con *vehículo*, *gato* ou *can*. Tamén son deste tipo as palabras que adoitan ter antónimos claros (*amigo-inimigo*). No caso dos adxectivos, son palabras que poden ir con outras de diferentes tipos no eixo sintagmático (*grande* podería acompañar a *casa*, *persoa*, *árbore*, etc). As non nucleares, ou expresivas, son as que soportan un valor connotativo que expresa actitudes ou emocións, aínda que os idiomas varían no modo mediante o que producen este efecto (*cachorriño*, *gatiño*). Poden acompañar a un menor número de tipos de palabras (*corpulento* en vez de *grande*) e resulta menos frecuente que se lles poda asociar un antónimo (*compañeiro* en vez de *amigo*).

Crystal (1995) introduce o termo *lexema* como abreviatura de *ítem lexical*, para referirse a tódalas unidades de significado lexical. Identifica tres tipos:

a) Palabras base ou raíz e as súas inflexións morfolóxicas e derivacións. O significado de cada palabra derivada depende do significado da palabra base (átomos sintácticos en terminoloxía de Pinker, 1994).

b) Palabras individuais, locucións e expresións en xerga. Son aquelas das o que o seu significado non é posible extraelo utilizando os regulamentos da morfoloxía. No caso de

locucións e no das expresión de xerga o seu significado non é o mesmo que o que resultaría da combinación literal das palabras individuais, teñen un significado coma se fosen unha soa palabra (listemas para Pinker).

c) Verbos multipalabra, que en inglés son aqueles acompañados dunha preposición e teñen un significado diferenciado do termo illado.

Pola súa parte, Goulden, Nation e Read (1990) clasifican as palabras que aparecen no *Webster's Third New International Dictionary*, considerado o máis amplo dicionario non histórico en inglés, nas seguintes categorías:

a) Palabras base: son as utilizadas como entradas principais nos dicionarios.

b) Derivadas: son as que requiren unha mínima aprendizaxe e é posible deducir claramente o seu significado a partir das súas compoñentes. Entre elas contabilizan as formadas con prefixos comúns.

c) Compostas: son aquelas formadas por dúas ou máis palabras separadas por un espazo ou guión.

d) *Proper words*: son as palabras que os dicionarios indican que usualmente, a miúdo ou ás veces se escriben con maiúscula.

e) Outras, entre as que se encadran símbolos, prefixos, sufixos, letras, abreviaturas, formas arcaicas e dialectais.

Nagy e Anderson (1984) utilizan un criterio semántico, ademais do morfolóxico, para diferenciar entre o que denominan palabras básicas (*basics words*) e palabras derivadas (*derived words*). Empregan para elo unha escala de seis niveis (de cero a cinco), que representan diferentes graos de posible relación semántica entre unha determinada palabra e as que a compoñen ou a palabra base a partir da que se forma mediante afixos. Consideran como palabras básicas aquelas que se poden encadrar dende o nivel 3 ata o 5 porque é dende o nivel 3 a partir do que se entende que son semanticamente opacas, é

dicir, non é posible deduci-lo significado das palabras que se sitúan neste nivel en función dos elementos que a compoñen, sen unha axuda importante do contexto.

Por outro lado, a investigación sobre a relación entre o incremento do vocabulario e a lectura en falantes nativos ten mostrado a importancia do coñecemento morfolóxico para a determinación de que palabras coñecerán os lectores (Graves, 1987), coñecemento que se desenvolve ata a adolescencia. En relación co desenvolvemento deste coñecemento morfolóxico está o concepto de **familia de palabras**, que fai referencia a un grupo de palabras asociadas entre si en razón dun elemento común (a raíz). Unha familia de palabras está constituída por unha palabra base e tódolos seus derivados e formas flexionadas que poidan ser encadradas sen necesidade de aprender cada forma por separado. É, polo tanto, un grupo de palabras cun étimo común ou vocábulo do que procede outro nunha lingua distinta (*demos* en grego é o étimo de *democracia*, *gramma* é o étimo de *gramática*). Os vocábulos que conforman unha familia de palabras gardan entre si unha relación non só de significado (como no campo semántico), senón tamén na forma, posto que teñen en común o lexema ou a raíz (Alonso, 1986). Nas palabras derivadas o significado esencial debe estar moi relacionado co significado da palabra base cando esta aparece soa ou con outras palabras derivadas (Bauer e Nation, 1993). Unha vez que se coñece a palabra base ou unha forma derivada, o recoñecemento doutros membros da familia presenta un escaso grao de dificultade.

Este tipo de conceptos (palabras derivadas, afixación, sufixación, familias de palabras, etc.) son cuestións que os autores de dicionarios teñen que considerar, xa que afectan ós termos a incluír, pero habitualmente trátanse de forma inconsistente tanto nun mesmo dicionario coma en dicionarios diferentes. Representan tamén unha cuestión metodolóxica a resolver en relación coa medida do tamaño do vocabulario, xa que ó dicir que unha persoa coñece un determinado número de palabras ou que un texto contén un determinado número de palabras debería estar claro qué se inclúe como unha palabra. Non

obstante, nas diferentes investigacións utilízanse criterios dispares neste sentido, de forma que segundo Bauer e Nation (1993), aínda que con matices distintos, autores como por exemplo Thorndike e Lorge (1924) ou West (1953) consideran os membros dunha familia coma unha soa palabra, mentres que outros como Carrol, Davies e Richman (1971) consideran coma unha palabra diferente calquera variante da palabra base. Sería conveniente, por elo, que se adoptaran criterios unificados co fin de poder comparar, ou cando menos describir claramente, que se considera como unidade/ítem a contabilizar.

Como se acaba de ver, non está claramente delimitado *que é unha palabra*, debido a que o que tradicional e comunmente se entende por palabra non é unha unidade lingüística coa que se poida explica-la relación entre significante e significado en tódalas situacións posibles. Polo tanto, parece máis útil entendela preferiblemente desde o punto de vista da forma que do significado, como “unha mínima forma libre”, como fai Bloomfield.

1.1.2 Léxico e vocabulario. Palabra e vocábulo

Lázaro Carreter (1981, p. 262) define o *léxico* como “o sistema de palabras que compoñen unha lingua”, denominándose lexemas as unidades que o integran. Rey-Debove (1976) exprésao dun modo distinto e posiblemente menos claro, pero coincidente. Para este autor o léxico sería “o conxunto de signos dunha lingua que pon en funcionamento a gramática (sintaxe/morfoloxía) desa lingua” (p. 167). Defínese “ben como o conxunto dos morfemas dunha lingua (léxico de Bloomfield), ben como o conxunto das palabras (mínima forma libre de Bloomfield) dunha lingua” (p. 167). Genouvrier e Peytard (1970, p. 181) introducen unha diferenciación entre léxico individual e léxico global ou xeral. Para estes autores “o léxico está conformado por tódalas palabras que están a disposición dun falante nun momento determinado. Son as palabras que pode empregar e comprender nese momento; constitúen o seu léxico individual”. Existe, ademais, un número indeterminado de palabras que un individuo en particular non tivo ocasión de ler ou escoitar antes, pero que

están a disposición dos membros da comunidade á que pertence, que constitúen o léxico global ou xeral: cada léxico individual non é máis que é unha parte de aquel.

O *vocabulario* pola súa parte remite ós lexemas que aparecen nunha ou varias mostras de lingua oral ou escrita dunha persoa, comunidade lingüística ou dunha ciencia (Justicia, 1995, p.11). Segundo Lázaro Carreter (1981, p. 411), o vocabulario é "o conxunto de palabras empregadas por un autor, un falante, unha escola literaria, unha ciencia, etc.". Genouvrier e Peytard (1970) definen o vocabulario nos seguintes termos:

O vocabulario é conxunto de palabras utilizadas por unha persoa nun acto concreto de fala. O vocabulario é a actualización no tempo de certo número de palabras pertencentes ó léxico individual do falante ou a actualización dunha serie de palabras que conforman o léxico da persoa. Neste sentido, é posible contabiliza-las palabras diferentes que unha persoa emprega nunha actividade de redacción ou expresión oral, o que achega unha idea da riqueza do seu vocabulario nese momento, pero non representa máis que unha mostra do seu léxico (p. 181).

Polo tanto, entre o vocabulario e o léxico existe unha relación de inclusión. O primeiro é unha parte, en dimensións variables segundo o momento e as necesidades, do léxico individual, que, á vez, o é do léxico global en terminoloxía de Genouvrier e Peytard. A diferenciación é semellante á establecida por Saussure entre lingua e fala (Martinet, 1972, p. 206-208) ou por Guillaume entre lingua e discurso, segundo Galisson e Coste (1976, p. 317). A lingua é entendida por Saussure como o repertorio utilizado por unha comunidade lingüística, que constitúe un conxunto de posibilidades que se lle ofrecen ós usuarios para que escollan entre elas co obxecto de construí-los enunciados que tratan de transmitir. A fala é entendida polo mesmo autor como enunciado ou conxunto de enunciados. O vocabulario sería a actualización de unidades virtuais, os lexemas, que forman o léxico. Neste sentido, o vocabulario presupón un léxico, do que é unha mostra, un subconxunto.

Para rematar este apartado, cabe facer referencia á diferenza entre dous conceptos que se relacionan cos de léxico e vocabulario (Justicia, 1995, p. 11) e que se empregarán máis adiante no punto 1.2.3 deste capítulo e no capítulo 3, cando se comente o concepto de vocabulario frecuente, así como algúns estudos sobre vocabulario. Un destes conceptos é o de vocábulo (*type*, tipo), que se utiliza para designar cada unha das unidades que compoñen o vocabulario dun suxeito. Son o resultado de liberar dos seus elementos gramaticais unha forma do discurso. Corresponderíanse con entradas que rexistran nun dicionario. O outro concepto é o de palabra (*token*, figuración), utilizado para designar toda ocorrencia de calquera palabra nun texto oral ou escrito. No caso do texto escrito, serían todas e cada unha das unidades elementais, perfectamente distinguidas pola tipografía que integran un texto escrito.

1.1.3 Palabra, oración, texto e situación

A palabra adquire significado pleno no contexto de unidades lingüísticas maiores como son a oración ou o texto, utilizadas para expresar aquilo a que se fai referencia, estando o seu significado determinado incluso pola función que o texto ten no contexto extralingüístico ou situación.

Segundo Luria (1978, p. 12; 1980, p. 47) as palabras están constituídas por dous elementos básicos: a referencia "obxectal" (función da palabra en canto á designación de obxectos, accións ou relacións) e o significado (a función de illar determinadas características esenciais do obxecto, de xeneralizar e incluír un obxecto nun determinado sistema de categorías). Sendo así, a xuízo de Luria (1980, pp. 35-41), pódese considerar que as palabras cumpren unha serie de funcións.

Unha delas é a designación, xa que as palabras denotan ou fan referencia, designan un obxecto, acción, etc., indican o obxecto correspondente no mundo externo (realmente o referente tamén pode ser algo interno, inmaterial, imaxinario). Outra función é a de análise

dos obxectos, xa que as palabras non soamente substitúen ós obxectos, designándoos, senón que permiten efectuar unha análise das súas características, illando os trazos esenciais dos menos esenciais, establecendo un sistema de complexos enlaces e relacións. Desto deriva unha terceira función, a de xeneralización do obxecto de referencia, en tanto que o inclúe nunha determinada categoría a partir da análise das características e do establecemento de relacións. As palabras tamén cumpren unha función de comunicación ou transmisión de información como consecuencia do anterior, posto que ó extrae-los trazos característicos e xeneraliza-lo obxecto, as palabras convértense en instrumentos do pensamento e en medios da comunicación. Deste modo, o pensamento existe a través das palabras: non se trata de que meramente se exprese mediante elas, xa que ó representar conceptos, configúrase como o aparello básico do pensamento (Vygotsky, 1962, pp. 125, 153). Finalmente, cumpren unha función de transmisión da experiencia acumulada dunha xeración ás posteriores.

En síntese, as palabras designan as cousas, individualizan as características das cousas, as accións, as relacións, organizan os obxectos en determinados sistemas. Dito doutro modo, codifican a experiencia das persoas xogando o papel de sistematizadoras da mesma, de modo que se podería dicir que practicamente constitúen esa experiencia en tanto que permiten representala de forma permanente. Para Luria esto permite que a persoa poida afondar e amplia-la propia percepción directa e conforma-la propia conciencia. A xuízo de Vygotsky (1962), as palabras, ademais de ser fundamentais no desenvolvemento do pensamento, tamén o son para o desenvolvemento da conciencia humana e representan “un microcosmos” desta conciencia (páx. 153). As citas que seguen son representativas da súa postura.

Xa que o significado da palabra (*word meaning*) é pensamento e fala, encontramos nel a unidade do pensamento verbal que estabamos buscando. Claramente, entón, o método a

seguir na exploración da natureza do pensamento verbal é a análise semántica – o estudo do desenvolvemento, o funcionamento, e a estrutura desta unidade, a cal contén pensamento e fala interrelacionados (p. 5).

O significado da palabra (*word meaning*) é un fenómeno do pensamento en tanto que o pensamento está expresado mediante a fala, e da fala en tanto que a fala está conectada co pensamento e iluminada por el. É un fenómeno de pensamento verbal, ou fala con significado – unha unión de palabra e pensamento (p. 120).

Por outra parte, hai palabras que cumpren a función de dotar de cohesión ó texto, unindo e relacionando adecuadamente unhas palabras con outras. Hasan (1985, p. 93), dentro do seu modelo de análise das relacións semánticas no texto, denominou *relevantes* a aquelas palabras que facilitan a cohesión do texto. Forman cadeas que se conectan entre si mediante diversos tipos de relacións semánticas. Unha destas relacións é a co-referencia: as palabras refírense á mesma cousa, evento, etc., como ocorre cos pronomes demostrativos, artigos definidos e comparativos; outro tipo de relación semántica é a co-clasificación, que inclúe a substitución e a elipse: as palabras refírense a eventos distintos da mesma clase (*Eu leo. O meu amigo faino tamén*, por exemplo); un terceiro tipo de relación semántica é a co-extensión: as palabras refírense a algo dentro do mesmo campo semántico (palabras sinónimas e antónimas, entre outras). As palabras relevantes, á súa vez, divídeas en centrais e non centrais. *Centrais* son as que facilitan a conexión de ideas interactuando de forma directa e recorrinte ó longo do texto, mentres que as *non centrais* son as que non manteñen unha relación de interacción directa. Finalmente, Hasan denomina *periféricas* ás palabras que non son esenciais para a coherencia nin para a cohesión do texto e que resultan da diferenza entre o número total de palabras nun texto e o número de palabras relevantes.

O coñecemento das palabras cumpre, asemade, unha importante función no proceso de comprensión dun determinado texto, fundamentalmente se o seu significado é esencial

no mesmo. Representan o que se pode considerar coma un primeiro chanzo porque permiten establecer relacións entre unhas e outras no que respecta ós seus significados, así como establecer inferencias e relacións co coñecemento previo sobre o que elaborar un grao de comprensión máis profundo.

Non obstante, como se indicaba anteriormente, a palabra adquire significado no contexto doutras unidades lingüísticas maiores. A xuízo de Vygotsky (1962, p. 151), para entender a outras persoas non é suficiente entende-las palabras que emite, é preciso entende-lo seu pensamento. Pero, aínda esto non é suficiente, xa que tamén se debe coñece-la súa motivación e esta motivación cabe entendela como a intención que predomina nun texto, qué trata de conseguir (informar, desinformar, divertir, convencer, persuadir, etc.). En palabras dun personaxe de *Alicia a través del espejo y lo que Alicia encontró al otro lado*, de Carroll:

- Cando emprego unha palabra, insistiu Zanco Panco con ton de voz máis ben desdeñoso, quere dicir o que eu quero que signifique ... nin máis nin menos.
- A cuestión, replicou Alicia, consiste en saber se podes facer que as palabras teñan tantos significados diferentes.
- A cuestión, dixo Zanco Panco, consiste en saber quen manda (1973, p. 116).

A palabra non parece que se poda considerar a unidade básica de significación, como xa se sinalou no punto 1.1.1 Si cabe planearse a oración como unidade básica de significación (Palmer, 1983, p. 37), porque se entende como a expresión dun pensamento completo e, se ben tanto as palabras coma as oracións teñen significado, o das palabras vese matizado ou complementado pola función que cumpren como parte da oración. Para clarificar esta idea, pénsese nunca situación en que unha persoa emite unha palabra inventada e o interlocutor interpela a esa persoa acerca do significado da mesma, o emisor,

para definila, seguramente utilizará oracións que expliquen o significado da palabra inventada.

Agora ben, convén non esquecer que hai significados que non están directamente relacionados coa estrutura gramatical (oración) nin coa lexical (palabra), nin sequera coa textual, senón que dependen de elementos prosódicos e paralingüísticos (entoación, acentuación, ritmo, etc.) ou expresións faciais e xestos, mecanismos que se utilizan, incluída a entoación, para indica-lo importante, o novo ou o contraste. Hai unha variedade de actos de fala (advertir, ameazar, prometer ...) que non é necesario expresar abertamente (*Espero que non o fagas, Hai un bicho en ...* non son meramente informativas). É posible dicir algo intentando transmitir algo diferente: por exemplo, dando información irrelevante, pódese transmiti-la idea de que realmente a información relevante non é favorable (se se fala da capacidade profesional dunha persoa e se di que *non obstante, é boa persoa*). Finalmente, as fórmulas sociais non significan realmente o que significan as palabras que as compoñen.

En resumo, as palabras en xeral organizan e sistematizan a experiencia, teñen gran importancia no proceso de comprensión de textos e algunhas contribúen á coherencia e cohesión. Non obstante, non se consideran como unha unidade básica de significación, senón que o significado está determinado en gran parte polo contexto en que aparecen (situación, texto, oración).

1.2 ¿Que significa coñecer unha palabra?

1.2.1 Unha primeira aproximación

De todo o exposto no punto anterior pode deducirse que non resulta tan sinxelo como en principio se puidera pensar responder á pregunta que se formula no epígrafe do apartado, o que se verá confirmado a medida que se trate de clarificar. Basta con cotexar nos dicionarios as definicións dos vocábulos *significar*, *significación* e *sentido* para decatarse

disto. Ogden e Richards (1923, en Palmer, 1983, p. 3) encontraron en inglés ata 16 significados diferentes do verbo *to mean* (significar). Outro exemplo móstrao Lyons (1980, p. 4), quen analiza diversos matices do término inglés *meaning* (significado), entre os que se encontran os de sentido, valor, querer dicir e intención.

Habitualmente establécese unha distinción entre significado e sentido, se ben tamén se confunden ámbolos dous conceptos con frecuencia. Luria (1980, p. 49) entende o **significado** como o sistema de relacións que comprende cada palabra, formado ó longo do proceso histórico; este sistema de relacións pode ser de distinta profundidade, distinto grao de xeneralización e distinta amplitude, pero sempre conserva un núcleo de significación relativamente estable, un determinado conxunto de enlaces que só se modifican polo contexto dentro duns límites. O **sentido** refírese ó significado da palabra nun contexto determinado; está composto por aqueles enlaces ou relacións que poidan establecerse nun momento e situación dados e recolle a achega dos aspectos subxectivos do significado nese momento e situación dados. Ten que ver co complexo sistema de relacións entre elementos lingüísticos (fundamentalmente palabras), ten que ver coas relacións intralingüísticas (Palmer, 1983, p. 29). O sentido queda establecido polos elementos presentes na cadea e polos ausentes, pero evocados por asociación. Tamén se pode entender coma o conxunto de representacións que suxire unha palabra ou coma o conxunto de representacións suxeridas por unha palabra nun caso determinado. O sentido dunha palabra resulta da suma de tódolos eventos que son evocados á nosa conciencia pola palabra e depende do contexto lingüístico da oración, o cal depende do parágrafo, que depende do texto e este doutros traballos do mesmo autor (Vygotsky, 1962, p. 146).

As palabras fan referencia (persoas, cousas, etc., determinadas como en *Este coche*) e denotan (clases de persoas, cousas, etc. como en *coche*), pero gran parte do seu significado ten que ver coa actitude do falante cara o referido, cara o que escoita e cara o propio acto de fala. Por elo (Yuen Ren Chao, 1975, p. 84), o significado nunha lingua é

moito máis complexo e interesante que o simple emprego dos símbolos para as cousas, nomes de xénero natural (casa, billa, can, etc.). O significado non debe ser entendido como algo inmutable, estático, senón como algo dinámico. Como afirman Harrison, Bayley e Dewar (1998):

O `significado' non é algo que poda ser considerado como inmutable. O `significado' das palabras non é fixo, porque o `significado' é un fenómeno tanto social coma lingüístico, como resultado do cal varía sutilmente en cada contexto de produción e interpretación (p. 6).

Segundo Vives Boix (1995, p. 47) ten sentido a noción de coñecemento do vocabulario como un contínuum, en tanto que as palabras non son só coñecidas ou descoñecidas, senón que tamén poden ser coñecidas parcialmente, que é o que realmente se ve moi claramente cos aprendices dunha segunda lingua, xa que lentamente van aprehendendo o propio significante e o significado da palabra. O coñecemento dunha palabra é “sempre un proceso activo de reconstrución” (Richards, 1976, p. 83), desenvólvese ó longo do tempo e as súas representacións son activadas continuamente, polo que o seu significado está conformado por nunha complexa rede de asociacións, matices, usos, palabras que normalmente coocorren con ela, etc. O desenvolvemento semántico dunha palabra, tanto a referencia ó obxecto coma a separación das correspondentes características, a codificación dos trazos dados e a inclusión do obxecto nun determinado sistema de categorías cambian ó longo do proceso evolutivo dunha persoa.

Existe un mínimo coñecemento dunha palabra incluso cando se encontra por primeira vez, xa que a competencia lingüística permite predicir razoablemente se se trata dunha clase ou outra de palabra, por exemplo. Realmente non quere decir que se coñeza o significado da palabra, senón que se recoñece como algo da lingua que se emprega. No caso dunha palabra illada consistiría en recoñecer que é unha palabra e non un son extralingüístico ou

unha palabra nunha lingua descoñecida. Esta capacidade pertence ó coñecemento do vocabulario, pero de carácter infrasemántico, xa que un suxeito pode basearse en criterios fonéticos cando a utiliza (Oleron, 1981, p. 70). Tamén se demostra certo coñecemento cando unha persoa recorda detalles dun significante (a primeira letra, número de sílabas, etc.) ou dun significado. Ademais, cando xa se produciron varios encontros cunha palabra, é necesario un menor número deles posteriormente para aprendela que cando se encontra por primeira vez.

Polo tanto, para considerar que se aprendeu unha palabra, cando menos, é preciso recoñecela como palabra e que a súa representación fónica forme parte do propio léxico mental. O carácter básico deste aspecto do coñecemento dunha palabra ponse de relevo nos traballos sobre asociacións verbais. Estes demostran que tanto escolares en lingua materna coma estudantes dunha segunda lingua producen en maior medida asociacións baseadas na representación fónica das palabras (asociacións tipo *clang*) que non adoitan producirse en adultos nativos nin en estudantes dunha segunda lingua de niveis intermedios ou avanzados, a non ser cando as palabras estímulo son pouco frecuentes ou se piden respostas rápidas (Beck, 1981; Clark, 1975, p. 287; Meara, 1978). Pero, para entender que se coñece unha palabra require, ademais, aprende-las súas propiedades sintácticas, o seu lugar na estrutura léxica (as relacións con outras palabras), as propiedades semánticas e asociacións que xera (sinónimos, antónimos, hiperónimos, hipónimos, polisémicas, palabras relacionadas, etc.), o concepto base que determina o seu lugar no sistema conceptual dunha persoa, así como o esquema das especificacións dos seus significados, o significado literal, as diferentes connotacións, as opcións morfolóxicas que ofrece e coñece-la súa forma falada e escrita (Bloom, 2002, p. 17; Ellis, 1995; Nagy e Scott, 2000; Nation, 1990, pp. 30-33; Schmitt, 1998).

Richards (1976), despois de resaltar que o vocabulario continúa ampliándose na idade adulta, mentres que o coñecemento sintáctico é comparativamente menor, sinalou

unha serie de supostos ou aspectos do coñecemento dunha palabra (Figura 1.1.), coa finalidade de orienta-la ensinanza do vocabulario, o desenvolvemento da competencia léxica. O feito de que os supostos fosen elaborados cunha intención didáctica, non para ser utilizados na avaliación do coñecemento das palabras, pode explicar que os dous últimos non aparezan antes, cando son de importancia capital en relación co coñecemento dunha palabra, así como que non aborde o coñecemento activo e pasivo do vocabulario ou que non faga alusión ás circunstancias en que son adquiridas as palabras. Non obstante, estes supostos, que en boa parte concordan cos aspectos citados no parágrafo anterior, foron utilizados por varios autores como marco para describir que significa coñecer unha palabra (Meara, 1996).

Coñecer unha palabra significa:

1. Coñecer o grao de probabilidade de encontrar esa palabra na fala ou escrita. Coñece-lo tipo de palabras ás que máis probablemente vai asociada ou que coocorren de forma habitual con ela (frecuencia, e, en relación con ela, a probabilidade de identificación como da propia lingua).
2. Coñece-las limitacións impostas no seu uso de acordo ás variacións en función do rol social, contexto social, etc. e da situación (uso pragmático).
3. Coñece-las propiedades sintácticas asociadas a esa palabra, fundamentalmente cando se trata de verbos.
4. Coñece-la vinculación da palabra base coas súas derivadas, é dicir, os membros da familia léxica.
5. Coñece-la rede de asociacións con outras palabras.
6. Coñece-lo valor semántico e as valoracións con que está "cargada" a palabra, os compoñentes semánticos (sentido e connotacións).
7. Coñecer moitos dos diferentes significados da palabra (sinonimia).

Figura 1.1.: Supostos do coñecemento dunha palabra (Richards, 1976)

Por outra parte, cada un dos anteriores supostos contén, á súa vez, novos supostos. Así, pódese considerar que trala quinta afirmación (“Coñece-la rede de asociacións con outras palabras”) hai varios supostos (Figura 1.2.)

- Cada palabra entra nunha rede de asociacións con outras.
- A rede que resulta tende a ser semellante para tódolos falantes dunha lingua.
- É posible especifica-la rede.
- Un falante determinado dunha lingua coñece a rede.
- A rede é fixa e estable.
- A rede é un trazo fundamental do léxico, unha propiedade da súa estrutura profunda.
- O léxico en bilingües non é significativamente diferente do dos monolingües.

Figura 1.2.: Supostos en que se sustenta a quinta afirmación de Richards (1976)

Na Figura 1.3. trátase de presentar de forma resumida as condicións relativas a diferentes niveis de coñecemento dunha palabra, atendendo ó significante e ó significado.

Significante

- Recoñece-la forma como unha palabra.
- Recoñece-la palabra total ou parcialmente.
- Recoñece-la forma como unha palabra da propia lingua.
- Recoñecela na súa forma oral e escrita.

Significado

- Coñecer palabras relacionadas (sinónimos, antónimos, familia léxica, coocorrencias, etc.).
- Coñecer diferentes acepcións.
- Coñece-lo significado connotativo.
- Coñece-lo seu uso pragmático.
- Coñece-las propiedades sintácticas.

Figura 1.3.: Resumo dos aspectos acerca do coñecemento dunha palabra

Meara (1996) entende que non cabe concibi-lo coñecemento das palabras como un contínuum, xa que realmente non está claro que dimensión sería esta no caso das palabras. Segundo este autor, o que se produce é un efecto tipo soleira na transición entre o vocabulario pasivo e o activo, aínda que puideran existir diferentes graos de “pasividade” e “actividade” dentro de cada unha das etapas.

As principais propostas para esta gradación son un nivel pasivo/receptivo e outro activo/productivo. Wesche e Paribakht (1996) desenvolveron unha *Escala de Coñecemento do Vocabulario* (*VKS* en inglés) con cinco niveis. A escala consiste basicamente en afirmacións que a persoa avaliada debe emitir sobre o seu coñecemento dunha palabra concreta. En base a estas afirmacións defínense os niveis de coñecemento da palabra (Figura 1.4.).

1. Non é familiar en absoluto.
2. É familiar, pero non coñece o significado.
3. Coñece un sinónimo correcto.
4. É usada de forma semanticamente adecuada nunha frase.
5. É usada nunha frase de forma semanticamente adecuada e de forma gramaticalmente correcta.

Figura 1.4.: Niveis de coñecemento dunha palabra na *VKS* (Wesche e Paribakht, 1996)

Non obstante, non está claro que realmente quede reflectida a progresión no coñecemento dunha palabra, ou cando menos non resulta fácil delimitar esta progresión. Así, non sería improbable que unha persoa fose capaz de utilizar unha palabra nunha frase sen coñece-lo significado, usando o contexto. Por outra parte, esta escala implica que unha vez alcanzado un determinado nivel, este grao de coñecemento é estable, cando seguramente non sexa así, xa que na práctica esquecémo-lo significado de palabras que en certo

momento puidemos utilizar; máis ben, prodúcese unha mobilidade dunha etapa a outra, entre o coñecemento e o esquecemento.

As palabras, dixemos, son como moedas, co significado e palabras asociadas nunha cara e os sons noutra. A fragilidade dos vínculos entre as dúas caras da moeda indica que o lexicón mental é unha mestura evolutiva. Este vínculo rómpese facilmente como mostra a común, pero frustrante experiencia de saber que existe unha palabra, estando bastante seguro do seu significado, pero sendo incapaz de darlle forma en sons, como cando un profesor di: ‘A metade do cerebro, pero non podo recorda-lo nome’. Máis tarde recorda o término *hemisferio cerebral* (Aitchison, 1987, p. 191).

Non parece moi acertado pensar que a comprensión dunha palabra sexa unha cuestión de todo ou nada. Os límites entre o significado das palabras é difuso, movedizo en función dos diferentes contextos, á vez que a capacidade de expresión implica non só que aumente o número de palabras do vocabulario, senón tamén saber como interaccionan entre si. Ademais, “as palabras non teñen un único significado, senón varios e o falante parece moverse entre eles sen excesivos problemas á hora de comprende-lo discurso” (Domínguez e Estévez, 1999, p. 159). Non obstante, os instrumentos que se adoitan utilizar na estimación do vocabulario semellan sustentarse na concepción de que o significado das palabras é estable e claramente delimitado, polo que, se ben é posible utilízalos para valoralo seu coñecemento, en tanto que a valoración das mesmas se efectúa de forma illada do contexto, non sempre é fiable debido ó comentado máis arriba no anterior parágrafo: prodúcese unha mobilidade no coñecemento das palabras; este coñecemento é variable no tempo.

Meara e Rodríguez Sánchez (1993) propoñen un modelo de catro etapas para valoralo coñecemento dunha palabra. A tarefa utilizada para valorar este coñecemento consiste en

solicitarlle á persoa avaliada que clasifique palabras nunha escala de catro categorías que se corresponden coas catro últimas etapas do modelo de Wesche e Paribakht (1996), porque consideran que eliminando a que reflicte total descoñecemento da palabra, a escala efectiva consta realmente de catro puntos; desta forma é posible valorar un número importante de palabras en pouco tempo. Utilizando un instrumento destas características, comprobaron que dúas semanas despois da primeira aplicación, ás mesmas persoas non sempre colocaban na mesma categoría determinadas palabras, fundamentalmente as menos frecuentes. Este modelo permítelles efectuar predicións sobre a evolución ou declive do vocabulario en termos globais durante un determinado período de tempo.

Como xa se sinalou máis arriba, é importante lembrar que ás palabras outórgaselles un significado referencial, que alude á relación entre a palabra e o obxecto, a noción ou a entidade á que se refire, pero tamén unha relación de sentido, que son as conexións de significado que mantén unha palabra con outras. As principais relacións de sentido son:

1. Sinonimia. Segundo Lyons (1980, p. 192) dúas ou máis expresións son sinónimas nun certo ámbito de enunciados soamente se son permutables nestes enunciados. Segundo Alonso (1986), dúas palabras son sinónimas cando teñen significante diferente, pero idéntico significado. Pódese dicir, polo tanto, que se refire á equivalencia en canto ó significado entre significantes diferentes ou que un mesmo significado realízase a través de significantes diferentes. Así e todo, moi poucas palabras son en realidade completamente sinónimas, no sentido de ser intercambiabiles en calquera contexto sen a máis leve variación do significado obxectivo ou subxectivo.

2. Polisemia. Existe polisemia cando un mesmo significante posúe diversos significados. Basicamente sería o concepto contrario á sinonimia, é dicir, cando unha soa palabra/forma ou sintagma está cargada de varios significados (algunhas teñen dous ou tres, pero, outras teñen incluso máis). Habitualmente utilízase como un procedemento para economizar, xa que se aproveitan varias veces o mesmo significante variando o significado,

aínda que sofre a posibilidade de xerar ambigüidade. O número de palabras polisémicas é considerable. Ó respecto, Johnson, Moe e Baumann (1983) encontraron que nun corpus de 9000 palabras había 6530 polisémicas, o que representa o 72%.

Ullmann (1972, pp. 180-189) cita cinco fontes a partir das que é posible xerar polisemia. Unha consiste en introducir trocos na aplicación. As palabras presentan certo número de matices diferentes segundo o contexto no que son usadas, de forma que a maioría dos sentidos xorden mediante cambios de aplicación, aínda que tamén pode operar outro factor que é o uso figurado (*atacar* pode referirse a encher, producir un efecto daniño ou figuradamente comer ata fartarse; *alto* pode referirse a tamaño, estatura ou posición nunha xerarquía ou nivel social; *ollo* pode referirse ó órgano, sentido da vista, oco nunha masa ou estar atento e vixiante).

Outra fonte é a especialización. As palabras teñen un significado xeral na linguaxe ordinaria e sentidos especializados en esferas máis restrinxidas (*interese* como actitude de ánimo ou como beneficio no ámbito de finanzas; *acción* como feito de actuar en xeral ou como partes dun capital ou unha intervención militar; *compañía* como acompañamento ou como grupo de actores ou tropa militar). A forma extrema de especialización prodúcese cando un nome común se converte nun nome propio que denota un só obxecto nun medio particular (Casa Branca, Igrexa, Comunidade, *City* en inglés, etc.).

A terceira fonte é a linguaxe figurada. Unha palabra pode recibir un ou máis sentidos figurados sen perde-lo significado orixinal, o vello e o novo permanecerán un ó lado do outro, mentres non haxa ningunha posibilidade de confusión entre eles (*ollo da lei*, *ollo do furacán*; *carreira pedestre*, *carreira profesional*). Neste grupo inclúense a metáfora como unha comparación condensada que radica na imaxinación do que fala ou escribe, a metonimia, que consiste en designar unha cousa co nome doutra coa que está relacionada (*ler a Ferrín* por le-las súas obras), entre outras figuras.

A cuarta fonte son os homónimos reinterpretados. Cando as palabras son idénticas en canto á forma e a diferenza de significados non é moi clara, tendemos a consideralas como unha soa palabra con dous sentidos. De tódolos modos, considérase moi pouco habitual este tipo de polisemia e os exemplos non son fáciles de encontrar.

A última fonte é a influencia estranxeira. Unha lingua pode influír en outra, entre outras formas, trocando o significado dunha palabra existente. O “préstamo semántico” como se denomina habitualmente, será particularmente frecuente onde haxa un contacto estreito entre dúas linguas e unha delas lle sirva de modelo á outra (en Estados Unidos a palabra española *ministro* adquire o sentido adicional de *clérigo protestante*, baixo a influencia da palabra inglesa *minister*; *introducir* significa tamén presentar por influencia de *to introduce* en inglés).

Das cinco fontes de polisemia, as tres primeiras (cambios de aplicación, especialización do significado e uso figurado) son as máis importantes e a reinterpretación de homónimos é a menos frecuente.

3. Homonimia. Os homónimos son palabras con idéntica forma fónica, homófonos, que se distinguen polo sentido, aínda que en moitos casos a homofonía vai acompañada de homografía (as palabras teñen unha forma gráfica idéntica). Débese diferenciar, polo tanto, dous tipos de homofonía, un derivado da evolución da forma de dúas palabras que parten de etimoloxías distintas para o mesmo significante (*chama* de lume e *chama* de chamar); o outro tipo, é cando se produce unha diverxencia semántica no significado dunha única palabra.

Para autores como Alonso (1986) ou Ullman (1972) os límites entre homonimia e polisemia son moi imprecisos, tanto que para Alonso son realmente o mesmo fenómeno. Certamente presentan semellanzas en tanto que nos dous fenómenos o falante dispón dun só significante e varios significados, pero na polisemia atribúenselle varias acepcións a unha única palabra, mentres que na homonimia distínguense varias palabras, como por exemplo

coma substantivo (procedente do latín e grego), *coma* do verbo comer (do latín), *coma* estado de inconsciencia (procedente do grego) e *coma* termo comparativo. Entre os criterios utilizados para diferencia-la homonimia e a polisemia adoitan citarse (Lyons, 1980, pp. 491-496) a etimoloxía, se ben pode darse o caso de que a orixe primixenia sexa a mesma; a afinidade de significado segundo o perciben os falantes nativos, aínda que non sempre hai coincidencia entre os mesmos e a non equivalencia sintáctica. Segundo Ullmann (1972, pp. 198-212), a homonimia pode xerarse de tres modos, aínda que o terceiro considérase moi secundario:

O primeiro denomínase converxencia fonética. É a causa máis común e consiste no desenvolvemento converxente de dous sons, de modo que dúas ou máis palabras inicialmente con formas diferentes coinciden na lingua falada e, á veces, na escrita (*pola*, contracción e *pola*, galiña nova; *chama* do verbo chamar e *chama*, lapa de lume). Tamén se pode dar pola perda das consoantes finais, como ocorre, por exemplo, en francés e inglés.

O segundo modo de xerarse homonimia é a diverxencia semántica. Unha única palabra con dous ou máis significados que chegan a separarse ata perde-la conexión e a polisemia dará paso á homonimia, quedando anulada a unidade da palabra (*gallo*, póla de árbore e *gallo*, parte de locución adverbial; *gravar*, representar algo nunha superficie e *gravar*, cargar cun imposto). Constitúe a réplica da reinterpretación de homónimos como unha soa palabra con dous sentidos.

A terceira maneira de xerarse é por influencia estranxeira. Hai palabras introducidas doutros idiomas que acadan o rango de homónimos. Unha vez establecida a palabra no novo idioma, adáptase ó sistema fonético local, de modo que pode chegar a coincidir con outras palabras da lingua receptora (practicamente é unha forma especial de converxencia de sons). A influencia tamén podería darse a través de préstamo semántico, pero é un proceso moi pouco frecuente. En calquera dos casos non parece que sexa un procedemento que se encontre con facilidade en galego.

4. Heteronimia. Segundo o dicionario da RAE (2001, p. 1204), a heteronimia é “un fenómeno polo que dúas palabras que corresponden a termos gramaticais en oposición proceden de raíces diferentes, como, por exemplo, *touro-vaca*”. Son heterónomos aqueles nomes que teñen unha forma para o masculino e outra para o feminino. A distinción de xénero non se fai por procedementos gramaticais, senón mediante a utilización de palabras distintas, como é o caso entre outros, amais do xa citado de *touro/vaca*, *home/muller*, *egua/cabalo*, etc.

5. Antonimia. Refírese á relación entre dous significantes con significacións contrarias como sucede no caso de *pequeno-grande*. Son palabras que pertencen á mesma categoría sintáctica e que se opoñen, dentro dunha clase semántica, por un ou varios trazos semánticos, sendo comúns o resto dos trazos (Galisson e Coste, 1976, p. 47). Considerouse xeralmente como o contrario da sinonimia, aínda cando a natureza das relacións semánticas que representan é moi diferente.

Lyons (1970, pp. 253-273) distingue varias posibilidades de sentido contrario ou oposición e atribúelles nomes diferentes para que a distinción sexa o máis clara posible. Unha das posibilidades é a complementariedade, como ocorre en *solteiro* e *casado*, *macho* e *femía*; a negación dunha das unidades léxicas implica a afirmación da outra, ó igual que a afirmación dunha implica a negación da outra. A segunda posibilidade de sentido contrario é a antonimia, caracterizada pola gradación: non vai acompañada dunha dobre implicación; así, “non é bo” non implica “é malo”. Un terceiro tipo de oposición é a inversión (*marido/esposa*; *maior/menor*; *pai/fillo*; *médico/paciente*), en que as ideas expresadas polos dous membros de cada par son idénticas. É equivalente dicir “Xosefina é nai de María” que dicir “María é filla de Xosefina”. Outros tipos de oposición léxica son a direccional (*arriba/abaixo*; *ir/vir*), a ortogonal (*verán/primavera/outono*; *norte/leste/oeste*) e antipodal (*verán/inverno*; *norte/sur*).

6. Hiponimia e hiperonimia. Poderían ser descritas como unha sinonimia asimétrica e refírense a unha relación de inclusión de determinadas palabras hipónimas respecto a outras hiperónimas. Por exemplo, *tulipán* e *rosa* son hipónimos respecto de *flor*; *casa*, *hospital*, *museo*, *escola* serían respecto a *edificio*. O dicionario da RAE (2001) define os hipónimos como as palabras cun significado incluído noutras (p. 1217) e ós hiperónimos como as palabras cun significado que inclúe a outras (p. 1216).

Como se acaba de ver ó longo deste apartado, non é fácil delimitar que significa coñecer unha palabra, porque no coñecemento dunha palabra están implicados diversos aspectos como son os diferentes graos de coñecemento que unha persoa poida ter desa palabra (partindo do recoñecemento como unha palabra do propio idioma) e as relacións de sentido da devandita palabra (sinonimia, polisemia, homonimia, antonimia, hiper e hiponimia). Estes aspectos precisan, matizan ese coñecemento da palabra. Unha vez efectuada unha primeira aproximación en torno á idea de que se considera que é coñecer unha palabra e da complexidade que entraña, cabe referirse á distinción que se adoita establecer entre diferentes tipos de vocabulario, en tanto que tamén é de utilidade para entender qué é unha palabra, á vez que se relaciona cos distintos graos en que unha persoa domina o coñecemento das palabras.

1.2.2 Vocabulario activo e vocabulario pasivo

Son termos que se utilizan en referencia á oposición establecida entre coñecemento activo, utilizado espontaneamente mediante a expresión verbal, e coñecemento pasivo, o que unha persoa é capaz de comprender e interpretar cando se lle presenta.

Pero hai cuestións en relación con esta distinción sobre as que non existe acordo, como é se se produce un continuum entre ámbolos dous tipos de vocabulario ou se é necesario o vocabulario pasivo para o desenvolvemento do activo, xa que o límite entre

ámbolos dous non está claro. Tampouco existe acordo acerca da utilización desta diferenza, xa que autores como Galisson e Coste (1976, p. 17) pensan que incluso é contraproducente na didáctica, mentres que outros, como Crystal (1995), entenden que cando se presentan datos sobre o tamaño do vocabulario dunha persoa, debería ofrecerse sempre información sobre o vocabulario activo e sobre o pasivo. Por outra parte, con frecuencia asimíase vocabulario activo a vocabulario frecuente e vocabulario pasivo a vocabulario dispoñible, cando o vocabulario dispoñible tamén é activo, do mesmo modo que o é o frecuente, xa que pode ser activado inmediatamente polo falante.

É comunmente aceptado (Aitchison, 1987; Chanell, 1988; Clark e Hecht, 1986; Laufer e Paribakht, 1998) que o vocabulario pasivo é máis amplo que o activo, pero a evidencia diverxe no relativo a canto maior é, e é escasa no que se refire a se un incremento do pasivo repercute automaticamente no do activo, se a diferenza entre ámbolos dous é estable ou se cambia ó longo do tempo. En canto á relación entre o tamaño dos dous tipos de vocabulario, segundo Chamberlain (1965) o vocabulario pasivo en nativos ingleses é cinco veces máis amplo que o activo. Benedict (1979) encontrou que os nenos comprenden ata 50 palabras no momento en que son capaces de producir as 10 primeiras palabras e estableceo que nos seis primeiros meses a razón da aprendizaxe do léxico pasivo era máis do dobre da razón do activo. Noutro estudo (Fenson et al., 1993, pp. 48-49) os datos mostraron que ós 12 meses a diferenza entre produción e comprensión era aproximadamente dunhas 10 a 80 palabras e ós 16 meses dunhas 65 a 190 palabras. Peraita (1986) valorou o coñecemento de 40 palabras con nenos e nenas de 2 a 3 anos, encontrando unha diferenza entre as puntuacións medias do vocabulario comprensivo e do expresivo de 10 puntos porcentuais a favor do comprensivo, pero tamén que existe unha relación entre eles, así como a posibilidade de que se produza unha transferencia entre ambos. Crystal (1995, p. 123) estudou a relación entre vocabulario pasivo e activo pedíndolles a tres suxeitos adultos (unha secretaria, un home de negocios e un catedrático) que indicaran se coñecían ben as palabras que se lles

presentaban, se as coñecían vagamente ou se non as coñecían; por outro lado, debían dicir se as usaban a miúdo, ocasionalmente ou nunca. Os resultados mostraron que dicían coñecer en torno a un 25% máis de palabras das que usaban na escritura ou fala. Tamén no coñecemento dunha segunda lingua se comprobou que o vocabulario pasivo é máis extenso que o activo (Michel, 1972).

En canto á relación entre vocabulario activo e pasivo, Laufer e Paribakht (1998) compararon o rendemento de escolares que cursaban inglés como segunda lingua en dúas condicións distintas en canto a tres dimensións do vocabulario: pasivo (comprensión do significado máis frecuente dunha palabra), activo controlado (as palabras que poden ser usadas se é necesario), activo libre (palabras que se poden elixir voluntariamente ou usar espontaneamente). En términos xerais, xa que se producen certas diferenzas entre os grupos de alumnado, prodúcese un desenvolvemento diferenciado, sendo máis lento e menos predicible o vocabulario activo; tamén parece que canto maiores son os vocabularios pasivo e activo controlado, tamén é maior o activo libre, á vez que é máis probable que sexan as palabras infrecuentes as que pasen ó vocabulario activo; ademais, o vocabulario pasivo está formado por un maior número de palabras frecuentes, se ben, canto máis amplo é, máis palabras infrecuentes inclúe, polo que o espazo entre os vocabularios pasivo e activo controlado é maior en palabras pouco frecuentes.

A valoración do vocabulario pasivo efectúase mediante diferentes procedementos. Un deles é a utilización de tests de elección múltiple, nos que se presenta unha palabra estímulo e unha serie de opcións, unha das cales reflicte o significado da palabra estímulo, debendo os suxeitos asocia-lo estímulo coa opción que consideran correcta; outros procedementos distintos son a declaración da persoa avaliada de que comprende as palabras que se lle pide, a achega de sinónimos ou dunha definición, a utilización da palabra nunha frase, a ilustración cun debuxo da palabra que se somete a valoración ou sinala-lo referente unha vez que o examinador o denomina. O dominio activo dunha palabra valórase observando se

aparece en mostras de lingua oral ou escrita, espontánea ou suxerida, pero tamén mediante a declaración da persoa de que a utiliza.

En síntese, por vocabulario activo enténdese aquel que se pode utilizar de forma espontánea na expresión verbal, mentres que por vocabulario pasivo enténdese o que é posible comprender, pero que non se emprega de forma espontánea. Por outra parte, adoita considerarse que o vocabulario pasivo é cuantitativamente maior que o activo, diverxendo as estimacións de diferentes autores respecto das proporcións que lle corresponden a cada un deles.

1.2.3 Vocabulario frecuente

A frecuencia refírese ao terreo da lingüística ó número de ocasións en que se produce unha unidade lingüística nun determinado discurso. Para Galisson e Coste (1976, p. 242) esta noción está unida ás nocións de utilidade, rendibilidade e dispoñibilidade, xa que canto máis frecuente é unha palabra máis rápida e espontaneamente pode ser activada, polo tanto máis dispoñible, no momento en que é necesario empregala. A frecuencia do vocabulario é un dos máis importantes aspectos sobre o funcionamento da linguaxe, en tanto que representa un retrato, aínda que parcial, de que palabras son usadas realmente de entre tódalas posibles. Hai abundante investigación que mostra que a frecuencia é un factor que inflúe na comprensión, produción e aprendizaxe da lingua (Suárez García, 1999).

Existen estudos sobre a frecuencia de palabras, tanto absoluta coma relativa, en diversos idiomas. Estes estudos, a grandes trazos, baséanse no recuento das palabras utilizadas en producións escritas e teñen como finalidade determina-lo vocabulario común dunha lingua mediante cálculos baseados en inferencias a partir da frecuencia de aparición das palabras nas devanditas producións. Estes recontos das frecuencias de aparición das palabras obtéñense utilizando a relación *type/token*, tamén denominadas correlativamente tipo/figuración ou vocábulo/palabra (punto 1.1.2). É dicir, utilizan a proporción entre o

número total de palabras, de formas, emitidas por unha mostra de suxeitos ou analizadas dun texto (*token*/figuración/palabra) e o número de vocábulos diferentes, unha vez descontadas as palabras que se repiten total ou parcialmente (*type*/tipo/vocábulo). As fontes para realiza-los recontos adoitan elixirse atendendo a criterios subxectivos ou ó principio de autoridade. No caso dos recontos de vocabulario infantil, o número total de palabras que serven como base, habitualmente establécese atendendo a unha serie de criterios: tendo en conta a información proporcionada por outras investigacións sobre a extensión do vocabulario da idade correspondente, do estrato social, etc.; a estabilidade da distribución de frecuencias, de modo que a continuación do reconto non modifique esencialmente as proporcións de vocabulario nas distintas frecuencias; que o coeficiente decrecente na orde de frecuencias sexa semellante ó doutro vocabulario coñecido; que se produza unha aceptable proporción entre as diferentes clases de vocábulos (nomes, adxectivos, verbos, etc.) e que a proporción entre o número total de palabras utilizadas como mostra e o número de vocábulos obtido, como índice da riqueza léxica e da frecuencia media dos vocábulos, permita consideralas como usuais.

En función da frecuencia de aparición dos vocábulos, elabóranse dicionarios de frecuencias que se utilizan tanto no deseño de investigacións como no deseño de programas de instrución ou na construción de tests. Os listados que compoñen estes dicionarios considéranse vocabularios fundamentais, básicos ou usuais porque son o resultado de recontos de vocabulario activo e o vocabulario de uso nunha determinada idade establécese en base á frecuencia con que ocorren os vocábulos. O dicionarios destas características presentan as entradas (*type*/tipo/vocábulos) atendendo a criterios de frecuencia e de orde alfabética. Algúns dicionarios de frecuencias citados por Smith (1997a) entre os existentes en inglés son: *A Basic Writing Vocabulary: The 10000 Most Frequently Usual Words* (Horn, 1926), *The Teacher's Word Book of 30000 Words* (Thorndike e Lorge, 1944) e *Basic Reading Vocabularies* (Harris e Jacobson, 1982). En francés, un dos existentes é o que Imbs

publicou en 1971 *Etudes statistiques sur le vocabulaire français. Dictionnaire de frequencies. Vocabulaire litteraire des XIX et XX siècles* (en Suárez García, 1999).

En español, entre outros, están publicados: *A graded Spanish Words Book* (Buchanan, 1927), *Recuento del vocabulario español* (Rodríguez Bou, 1952), *Vocabulario usual, vocabulario común y vocabulario fundamental* (García Hoz, 1953), *Frequency Dictionary of Spanish Words* (Juilland e Chang-Rodríguez, 1964), *Vocabulario del español hablado* (Márquez Villegas, 1975), *El vocabulario general de orientación científica y sus estratos* (García Hoz, 1976b), *Vocabulario básico del español y sus aplicaciones a la enseñanza* (Díaz Castañón, 1977), *Vocabulario básico en la E.G.B.* (Casanova e Rivera González, 1989), *Diccionario de frecuencias de las unidades lingüísticas del castellano*, (Alameda e Cuetos, 1995), *El desarrollo del vocabulario: diccionario de frecuencias* (Justicia, 1995), *LEXESP. Léxico informatizado del español* (Sebastián, Martí, Cuetos e Carreiras, 2000), *La adquisición del lenguaje* (Serra, Serrat, Solé, Bel e Aparici, 2000).

A continuación coméntanse con maior profundidade algúns dos publicados en español e galego. Inclúese un comentario máis amplo dos mesmos porque constitúe un contido da contribución empírica que habitualmente está disperso e porque pode ser de utilidade como complemento ó que se exporá no capítulo 2 en relación cos estudos de vocabulario que tratan de determina-lo vocabulario activo coñecido.

1. *Vocabulario usual, vocabulario común y vocabulario fundamental* (García Hoz, 1953)

O autor utilizou catro tipos de materiais para o reconto de palabras que se corresponden con catro ámbitos distintos: vida familiar (cartas), vida social indiferenciada (periódicos), vida social regulada (documentos oficiais, BOE, boletíns eclesiásticos e sindicais) e vida cultural (libros). Contabilizou ata 100.000 palabras de cada fonte, debido a que a partir deste número a posibilidade de que aparecesen palabras diferentes era moi reducida (menos de 1 cada 90 palabras), polo que dificilmente se poderían considerar como

usuais estas palabras. O reconto final formárono 400.000 palabras (100.000 X 4), das que 12.913 son palabras non repetidas, cifra á que se se lle suman os vocábulos específicos de ámbitos profesionais, de lugares, etc. que cada persoa poda coñecer, aproxímase á cifra de 15.000 que adopta como referencia, seguindo a Watts (1946).

O material que usou, como se mencionou máis arriba, en primeiro lugar, consistiu en 620 cartas de persoas pertencentes a “clases humildes” de 15 provincias do estado español, das que obtivo 100.000 palabras. En segundo lugar, empregou 25 dos periódicos de maior difusión de distintas zonas publicados entre xaneiro de 1943 e decembro de 1948, de diferentes épocas para variar os temas tratados (500 artigos doutrinais, 2000 de noticias, 1000 de espectáculos e 500 de anuncios); de cada periódico recompilou 4000 palabras co que en total recolleu 100.000 palabras. En terceiro lugar, empregou documentos oficiais (12 de BOEs de distintos meses, 9 eclesiásticos e 14 sindicais), recompilando 33.333 palabras de cada tipo de documento. Finalmente, usou libros elixidos pola súa difusión desde 1943, segundo a información proporcionada por 10 editoriais, recollendo as 100.000 palabras de nove libros de adultos e de 13 libros de literatura infantil.

Para elaboralo listado de frecuencias, rexistrou coma unha soa palabra os significantes con máis dun significado. No caso de variacións debidas a accidentes gramaticais, en nomes, adxectivos e pronomes non persoais, rexistrou a forma en masculino singular e en verbos a forma en infinitivo. As palabras compostas rexistrounas como unha ou dúas, segundo os casos, e as que non aparecen no dicionario da RAE excluíunas do vocabulario usual. As palabras técnicas ou que denominan ciencias que non aparecen no dicionario da RAE tamén as excluíu cando presentaban unicamente frecuencia 1, incluíndoas se presentaban máis dunha frecuencia. Deste modo, elaborou vocabularios nos que figuran as frecuencias de cada palabra en cada un dos catro materiais utilizados e a frecuencia total de cada palabra. Un dos vocabularios é o *Vocabulario usual*, constituído por 12.428 palabras que figuran como de significado xeral no *Diccionario Manual de la Lengua Española* da RAE

(1950). De forma complementaria aparece unha relación de 253 palabras que figuran no dicionario como de significación restrinxida ou especial e outra relación de palabras que non aparecen no dicionario. As palabras que figuran en tódolos materiais empregados constitúen o *Vocabulario común* (1971 palabras e 339.110 frecuencias). O autor ofrece tamén unha relación de 212 palabras que non aparecen en tódolos materiais, das que 97 teñen máis de 40 frecuencias e 121 superan ese número de frecuencias ó suma-las formas dun vocábulo (*acentuar* e *acento*, por exemplo). Por último, elaborou un *Vocabulario fundamental*, constituído por 208 palabras con frecuencias distribuídas practicamente en partes iguais nos diferentes tipos de materiais.

2. *Frequency Dictionary of Spanish Words* (Juilland e Chang-Rodríguez, 1964)

Recompilaron un corpus de 500.000 palabras, entre as que seleccionaron un vocabulario básico de 5024 palabras, cun coeficiente mínimo de uso de 3,08 (concepto que se explica máis abaixo).

O dicionario consta de dúas partes. Na primeira proporcionan información sobre as 5024 palabras/entradas (*hermoso*, por exemplo) e as variantes (*hermosa*, *hermosas*, *hermoso*, *hermosos*) en orde alfabética. Non inclúen variantes da entrada no caso de palabras como son os adverbios ou, cando sendo palabras con posibilidades de variantes, aparece no universo léxico unicamente na forma que figura como entrada do dicionario. Cada entrada vai seguida do seu coeficiente de uso (resultado de multiplica-la frecuencia total polo coeficiente de dispersión e de dividir por 100), o coeficiente de frecuencia (suma de tódalas frecuencias parciais das distintas fontes) e coeficiente de dispersión (calculado a partir das diferenzas entre as frecuencias das distintas fontes), ademais do número total de ocorrencias de tódalas variantes da palabra, así como a frecuencia nos diferentes tipos de fontes utilizadas (obras de teatro, novelas, ensaios, artigos periodísticos e obras técnicas).

Cada unha das variantes da entrada vai seguida da frecuencia total desa variante e a súa frecuencia nos distintos tipos de fontes.

Na segunda parte, os autores proporcionan información en tres listas paralelas sobre as 5024 entradas en orde decrecente de coeficiente de uso na primeira columna, de frecuencia na columna central e de dispersión na terceira columna. As tres listas están divididas en 10 grupos de 500 palabras cada un. Cada palabra vai seguida polo correspondente coeficiente.

3. *El vocabulario general de orientación científica y sus estratos* (García Hoz, 1976b)

García Hoz utilizou o libro de texto considerado máis representativo de cada unha das doce materias dos estudos de bacharelato, agás de idiomas clásicos e modernos. Para o reconto baseouse nos criterios da RAE, rexistrando como vocábulo soamente as palabras base non flexionadas, mentres que as compostas, os xiros e frases contabilizounas como unha soa, excluíu os signos numéricos e recolleu algúns latinismos. O corpus consta de 2.302.360 frecuencias (palabras rexistradas ou *tokens*) das que 25.402 son palabras diferentes ou *types*. Para a súa organización utilizou o criterio de repartición (número de materias diferentes en que aparece), resultando os estratos *Vocabulario común de materias* (palabras que aparecen en tódalas materias) e *Vocabularios específicos* (palabras que aparecen só nunha das materias). Entre estes dous vocabularios situou o *Vocabulario común de áreas* composto polas palabras que aparecen nas diferentes áreas (lingüística, matemática, ciencias da natureza, social e antropolóxica) e os *Vocabularios compartidos* compostos por palabras que aparecen en diferentes materias, pero non en todas.

Utilizando o criterio da frecuencia (número de veces que aparece unha palabra no texto), estableceu estratos en función da mesma (palabras con frecuencia 1, frecuencia 2, etc.). Máis da metade das palabras teñen unha frecuencia relativa inferior a 1, o que lle permitiu establecer un estrato coas palabras máis frecuentes (as que teñen maior

probabilidade de aparecer unha vez cando menos en cada 100.000 palabras en vez de nas 2.302.360 reais) e as menos frecuentes (non alcanzan esta probabilidade). Obtivo o *Vocabulario general de mayor frecuencia* (6100 palabras). Das 25.402 palabras do *Vocabulario general de orientación científica*, cada termo vai acompañado dun número que indica a frecuencia relativa (obtida considerando como se se recontaran 100.000 palabras de cada materia), co fin de unifica-los datos de todas, xa que o número de frecuencias son diferentes en cada unha delas. O *Vocabulario común de materias* está constituído por 587 palabras e no listado aparecen para cada termo a frecuencia absoluta, a relativa, a porcentaxe que representa ese término no total dos términos, a porcentaxe que representa o término en cada materia e a suma das frecuencias relativas de tódolos terminos (para unha explicación máis completa ver pp. 169-170 do texto orixinal).

O *Vocabulario común de áreas* está formado por 2274 palabras que cobren o 83% das frecuencias relativas totais do *Vocabulario general de orientación científica*. O *Vocabulario general de mayor frecuencia* recolle as palabras cunha frecuencia relativa superior a 1; son 6100 palabras, o 92,52% das frecuencias relativas totais do *Vocabulario general de orientación científica*. A posibilidade de que algunha destas palabras apareza nun texto científico é de 0,9252. O 48,3% das palabras léxicas do *Vocabulario general de orientación científica* son comúns a máis dunha materia.

4. *Vocabulario básico del español y sus aplicaciones a la enseñanza* (Díaz Castañón, 1977)

Os autores utilizaron unha mostra de escolares entre 8 e 17 anos de Asturias. Enviaron impresos a centros para a recollida das palabras, que consistían nunha enquisa libre (escribir en 5 minutos tódalas palabras que fose posible) e nunha enquisa controlada, consistente en escribir as vinte primeiras palabras que se lle ocorreran sobre temas familiares (animais, campo, casa e mobles, cidade, comidas e bebidas, oficios, escola e material escolar, xogos e deportes, medios de transporte, partes do corpo e vestidos).

Recolleron un corpus de 85.000 palabras, cunha media de 53 palabras por escolar. No listado figuran as frecuencias por idades (8-9 anos, 10-11 anos, 12-13 anos, 15-16 anos e 17 anos). Os vocábulos foron incluídos en base ós criterios que seguen: cada vocábulo aparece coma unha entrada diferente; as flexións dunha palabra unífanse nunha soa entrada, pero figuran as flexións na mesma; os sinónimos aparecen coma entradas independentes; os homónimos e homógrafos figuran nunha única entrada, sinalando debaixo as variantes; recóllense neoloxismos aceptados pola RAE; os vocábulos base considéranos coma entradas á parte. En base á enquisa libre elaboraron unha lista alfabética de tódolos vocábulos recollidos (9782), unha lista por orde de frecuencia decrecente ata chegar á frecuencia 20 (1350 vocábulos) e, utilizando esta última lista, elaboraron outra por orde alfabética indicando o número de orde que lle corresponde a cada vocábulo. Coa enquisa controlada, confeccionaron unha lista alfabética cos vocábulos de cada tema e unha lista de frecuencia decrecente ata chegar á frecuencia 20; elaboraron tamén dúas listas conxuntas de tódolos temas suxeridos, unha delas cos vocábulos ata a frecuencia 20 e outra lista por orde alfabética con indicación do número de orde de cada vocábulo.

5. *Vocabulario básico en la E.G.B.* (Casanova e Rivera González, 1989)

No punto 2.3.3 explicaranse con maior amplitude determinados aspectos da investigación, como as probas que administraron e os resultados que obtiveron. Aquí unicamente se comentará de forma breve como están organizadas as diferentes partes do traballo. No primeiro volume presentan un vocabulario tipo, que está composto por 10.017 vocábulos, os cales foron extraídos do dicionario da RAE (1970) mediante o sistema de xuíces, en función da dificultade que puideran supoñer para a súa aprendizaxe. As autoras presentan os vocábulos por ciclos de EXB (1342 de Ciclo Inicial, 3940 de Ciclo Medio e 5185 do Ciclo Superior), indicando a acepción ou acepcións apropiadas para cada un e agrupados en 20 campos nocionais, dentro de cada un dos cales os vocábulos van ordenados

alfabeticamente. Para a elaboración do vocabulario tipo prescindiron de palabras gramaticais, incluíron formas derivadas cando supoñen un cambio de categoría gramatical respecto da palabras base (verbo/nome ...) ou de natureza semántica (concreto/abstracto...), incluíron na mesma entrada a forma en masculino e feminino de termos de parentesco e de nomes de animais, aínda que no dicionario aparezan dúas entradas e figura a forma en feminino de nomes de oficios e profesións, aínda que no dicionario non figure. Nun segundo momento efectuaron un reaxuste con suxestións de mestres e outros especialistas, á vez que confrontaron as entradas coas existentes en diferentes vocabularios (García Hoz, ICE da Universidade de Oviedo, Juilland e Chang-Rodríguez, Márquez Villegas, Rodríguez Bou). Ademais, presentan unha proba de vocabulario elaborada cunha mostra dos vocábulos básicos de cada ciclo. Dos resultados da aplicación destas probas a escolares de EXB españois, as autoras conclúen que a comprensión do vocabulario básico proposto é realmente alta. Por outra parte, inclúen neste volume unha serie de recomendacións para interpretar adecuadamente o cadro do vocabulario tipo de cada ciclo. Nun segundo volume inclúen 18.881 palabras e expresións diferentes extraídas entre as emitidas por escrito pola mesma mostra de escolares. As autoras ofrecen este vocabulario activo clasificado en tres listas: ordenado alfabeticamente, distribuído por comunidades autónomas e clasificado por frecuencia de maior a menor, así como unha comparación do vocabulario tipo e o activo, ordenado alfabeticamente e facendo referencia ós tres ciclos de EXB. Coa comparación entre o vocabulario activo e o vocabulario tipo, observaron que os escolares utilizaban o 71% do vocabulario tipo. No terceiro tomo recollen tres experiencias de ensinanza de vocabulario.

6. *Diccionario de frecuencias de las unidades lingüísticas del castellano* (Alameda e Cuetos, 1995)

Alameda e Cuetos partiron dun corpus 2.000.000 de palabras extraídas de 606 textos escritos pertencentes a diferentes xéneros (novela, ensaio, periodístico e científico),

publicados entre 1978 e 1993. O 50% proceden de novelas, o 25% de periódicos, o 15% de ensaios e o 10% de publicacións científicas. Os textos latinoamericanos representan o 10% do corpus.

O dicionario consta de dous volumes e inclúe tamén información sobre a frecuencia de letras, bigramas e sílabas. O primeiro volume comprende unha lista ordenada alfabeticamente das diferentes unidades lingüísticas acompañadas do número de ocorrencias no corpus. No segundo volume figura unha lista de tódalas unidades e a súa frecuencia de maior a menor. Os homónimos non aparecen en entradas diferentes, o que implica que as palabras polisémicas anótanse cada vez que aparecen sen ningunha especificación do seu significado. As formas dos nomes segundo xénero e número, así como as diferentes persoas dos verbos, preséntanse coa súa frecuencia relativa.

7. El desarrollo del vocabulario: diccionario de frecuencias (Justicia, 1995)

Este autor utilizou unha mostra de 3402 escolares entre 6 e 13 anos das provincias de Almería, Granada, Xaén e Málaga. Partiu dun corpus de 528.544 palabras, recollidas de mostras escritas (redaccións) e asociacións libres. Utilizou dous criterios para organiza-la estrutura e para determina-lo número de palabras: conseguir un número de palabras que se aproximase ó considerado como termo medio para suxeitos das idades estudadas e obter unha frecuencia media no conxunto das palabras que fose suficiente como para podelas considerar usuais.

O vocabulario quedou constituído por 8937 vocábulos e o dicionario está organizado en catro partes:

- 1ª. Listado de vocábulos ordenados alfabéticamente (vocabulario usual).
- 2ª. Listado de vocábulos por orde de frecuencia (vocabulario usual).
- 3ª. Vocabulario común formado por 2123 vocábulos comúns a tódalas idades.

- 4ª. Dicionario silábico constituído por un listado de sílabas por orde alfabética e outro listado de sílabas por orde de frecuencia.

8. *LEXESP. Léxico informatizado del español* (Sebastián, Martí, Cuetos e Carreiras, 2000)

O corpus consta de 5.020.930 palabras extraídas de textos narrativos, de divulgación científica, de ensaio, prensa, semanarios e prensa deportiva. Para a recollida de palabras, seleccionaron textos ó azar desde 1978 a 1995, excluindo literatura infantil, textos poéticos, teatro e traducións.

É unha base de datos, de palabras en español, editada en CD-Rom e que se compón de dúas partes. A primeira utiliza o corpus como base e permite realizar buscas de aparición e coaparición de palabras, de lemas de palabras, de categorías (nome e verbo) de palabras dentro dun contexto, así como o número de veces que unha palabra, lema ou categoría aparece no corpus e a frecuencia de aparición entre un millón de palabras.

A segunda parte ofrece a posibilidade de facer estudos que permiten obter listados de palabras atendendo a características determinadas, incorporando tres estudos lingüísticos deste tipo. Para a súa elaboración seleccionaron as 6500 palabras de maior frecuencia que foran nomes e verbos en infinitivo. Dos nomes incluíron as variacións de xénero e número cunha frecuencia maior a 6. No relativo ós verbos, incluíron, ademais as formas reflexivas. A partir destas palabras construíron cuestionarios que lles pasaron a estudantes universitarios e cos datos obtidos calcularon índices de concreción, familiaridade e imaxinabilidade.

9. *La adquisición del lenguaje* (Serra, Serrat, Solé, Bel e Aparici, 2000)

O apéndice desta obra contén un listado con información sobre as frecuencias de palabras utilizadas en lingua oral por dez nenos e nenas de entre un e catro anos. As mostras foron obtidas mediante gravación magnetofónica. O listado está organizado por orde de frecuencia decrecente e dividido en tres categorías: de 12 a 23 meses, 24 a 35

meses e 36 a 47 meses. Para o rango de 12 a 23 meses figuran 734 vocábulos e 7523 ocorrencias de palabras; para o período entre 24 e 35 meses, 3447 vocábulos e 58.197 palabras e entre os 36 e os 47 meses, 5246 vocábulos e 105.577 palabras.

En galego hai publicados diferentes tipos de vocabularios e glosarios. Entre estas obras cabe citar algunhas que se cinguen a ámbitos específicos de coñecemento como son *Glosario de orientación escolar e profesional* (Sobrado e Porto, 2000), que contén 729 termos en galego, inglés, francés e alemán, *Léxico de métodos de investigación e diagnóstico en educación* (Rubal, 2002), composto por 1907 termos en galego e castelán ou *Vocabulario xurídico administrativo galego-castelán* (Gómez e Torres, 2005), que recolle 1459 palabras en galego e español. Tamén existen vocabularios básicos de carácter xeral. Un exemplo destes últimos é o *Vocabulario básico da lingua galega* de Asaka (1996), que contén arredor de 3000 entradas e consta de tres partes. Na primeira a forma galega vai acompañada da correspondente en castelán e en xaponés, ofrecendo, ademais, información gramatical. Na segunda, en primeiro lugar, aparece a palabra en español e ó lado a forma en galego. Na terceira, a palabra en xaponés é acompañada da forma en galego.

Outro traballo relacionado cos que se acaban de describir é *Nivel Soleira* (Fernández, Monteagudo e Puga, 1993). É tamén un vocabulario básico máis que un dicionario de frecuencias, aínda que pretende recolle-los términos máis frecuentemente utilizados. Segundo os autores, non está baseado en estudos rigorosos, xa que seleccionaron as palabras baseándose na intuición e experiencia como falantes nativos en galego, polo que ten un carácter un tanto arbitrario. O contido trata de aborda-los termos que se poden considerar como os necesarios para a meirande parte dos aprendices de galego e abarca-lo nivel mínimo para a aprendizaxe, se ben, a xuízo dos autores, non resulta doado establecer este nivel. Recolle tanto expresións coma palabras de catorce ámbitos (servizos, lugares de lecer, etc.), oito temas (casa e fogar, comidas e bebidas, etc.) e tres formas lingüísticas

(actos de fala, nocións xerais e nocións específicas). Está deseñado esencialmente para adultos residentes en Galicia que viven e traballan en zonas urbanas, ben castelanfalantes que queren aprender galego, ben galegofalantes que teñen contacto coa lingua falada, pero non dominan activamente a norma estándar e desexan acceder a ela, sen descarta-la posibilidade de que permita unha adaptación ó ensino primario ou secundario.

Existen, así mesmo, unha serie de corpus léxicos informatizados elaborados polo Seminario de Lingüística Informática da Universidade de Vigo, coa dirección de X. Gómez (consultar en <http://sli.uvigo.es>). Un destes corpus é o *Corpus Técnico do Galego* (2003-2006) que permite pescuda-lo número de veces que aparece unha determinada palabra/forma en textos de distintos ámbitos. Á súa vez, está formado polo Corpus *GALEX* (2.606.400 palabras de textos xurídicos e administrativos), Corpus *XIGA* (2.100.516 palabras de informática e telecomunicacións), Corpus *AUGA* (1.881.562 palabras de ecoloxía e ciencias medioambientais), Corpus *ACHEGA* (1.531.039 palabras de economía) e Corpus *MEGA* (921.575 palabras de textos xornalísticos). O Centro Ramón Piñeiro para a Investigación e as Humanidades tamén elaborou un corpus con dúas versións almacenado en formato electrónico, denominado *CORGA* (*Corpus de referencia do galego actual*) e que está composto por textos de diversa procedencia que foron seleccionados segundo tres criterios (medio, data e área temática). A primeira versión consta de 17.646.907 *tokens* e 329.803 *types* e a segunda, que aínda non está totalmente rematada, consta de 13,3 millóns de *tokens* e 266.282 *types*. A primeira versión está elaborada a partir de libros (66,81%), xornais (26,41%) e revistas (6,78%) con datas desde 1975 a 2004. Os textos abarcan as áreas de economía e política (20,67%), cultura e artes (7,06%), ciencias sociais (12,33%), ciencias e tecnoloxía (6,49%), ficción (43,49%) e outros (9,98%), onde se inclúen deporte, turismo, etc. Amais, contén unha relación das mil formas máis frecuentes. Este corpus permite pescudar de cada forma a frecuencia con que aparece por data, por medio e por área temática.

Como dicionario de frecuencias en sentido estrito, M^a. C. Teijido, M^a. Luz Pérez e I. Fraga (2001) publicaron o *Diccionario de frecuencias léxicas do galego escrito para nenos. LEXIGAL-Primaria*. Está editado en CD e foi elaborado a partir de tódolos libros escritos orixinalmente en galego, publicados entre 1985 e 1999 e dirixidos a nenos e nenas de cero a doce anos. Estes libros pertencían ós xéneros de narrativa, poesía, teatro e cómic. Incluíron, ademais, textos escolares de Lingua Galega e Coñecemento do Medio de primeiro a sexto de Educación Primaria.

No caso dos dirixidos a nenos e nenas entre cero e cinco anos, utilizaron os textos completos, mentres que no dos dirixidos a nenos e nenas entre seis e doce anos utilizaron páxinas ó chou. Deste modo, recompilaron seis mil palabras de media, se ben, como advirten as autoras, non todas se corresponden coas aceptadas pola normativa, xa que se recolleron tal como aparecían nos textos. As cantidades de palabras recompiladas segundo o tramo de idade son as seguintes: entre cero e cinco anos, 14.711; entre seis e sete anos, 23.966; entre oito e dez, 60.329 e entre once e doce, 53.718 palabras. Cada palabra acompáñase da súa frecuencia absoluta (número de veces que aparece) e a ponderada (número de veces que aparecería se a mostra fose de 1 millón). O sistema posibilita a busca de palabras que comecen, rematen, sexan iguais ou diferentes á letra, conxunto de letras ou palabra que se solicite, á vez que selecciona-lo tramo ou tramos de idade que interesen, de modo que, efectuada a busca, aparece a palabra ou palabras solicitadas e as frecuencias por tramo de idade.

En resumo, a frecuencia do vocabulario estímase en base á relación entre o número total de formas/palabras que aparecen nunha mostra de expresión oral ou escrita e o número de formas/palabras diferentes (vocábulos) que contén esa mostra de linguaxe que resultan unha vez eliminadas as formas que se repiten ou que son variacións da forma base. A partir dos recontos, elabóranse dicionarios de frecuencias que permiten consulta-la

frecuencia de aparición dunha determinada palabra. A frecuencia dunha palabra constitúe unha variable de suma importancia no que se refire ás posibilidades de que esa palabra sexa coñecida, tanto a súa forma (significante) como o seu significado: as palabras que aparecen máis veces nos textos, coas que, polo tanto, é máis probable que se teña maior experiencia, recoñécense máis pronto que aquelas outras de baixa frecuencia, á vez que existen maiores probabilidades de que sexa coñecido o seu significado.

1.2.4 Vocabulario dispoñible

A xuízo de López Morales (1984, p. 62) “o vocabulario dispoñible reflicte o caudal léxico utilizable nunha situación de comunicación determinada”, “o vocabulario concreto que o falante podería utilizar nun contexto comunicativo dado” (p. 63). Segundo o citado autor, examinando listas de frecuencias obsérvase que hai palabras cunha moi baixa frecuencia ou que nin aparecen nas listas e que, non obstante, a impresión é de que parecen ser amplamente coñecidas e utilizadas. Por esta razón, é necesario non confundir o que é unha situación frecuente co que é léxico frecuente, nin este co que é léxico fundamental, xa que hai palabras que só son utilizadas ó falar ou escribir de temas moi concretos, mentres outras aparecen en conversas ou escritos sobre calquera que sexa o tema. Para recolle-lo vocabulario dispoñible pero non frecuente, utilízanse probas de asociación en torno a “centros de interese” ou temas. O vocabulario rexistrado representa aquel de que dispón unha persoa realmente cando a súa expresión se cingue a eses temas. Para a obtención do número de vocábulos, límítase o tempo de realización ou péchanse as listas tras un número de palabras previamente determinado. Este concepto de vocabulario dispoñible está en relación co vocabulario usual e coa posibilidade de utilización inmediata polo falante, segundo as necesidades de produción lingüística. Está composto fundamentalmente por nomes, que son palabras cunha frecuencia menos estable, é dicir, son as palabras con menos posibilidades de aparecer en moitos contextos diferentes (un determinado verbo pode

aparecer en diversos contextos e o mesmo sucede con adxectivos, mentres que un nome en particular ten menos posibilidades de aparecer en contextos distintos) o que non ocorre nos listados de frecuencias, nos que abundan máis as palabras gramaticais, seguidas de verbos e adxectivos (p. 63).

1.2.5 Vocabulario usual

Para García Hoz (1977) o vocabulario usual é asimilable ó que se considera corrente, vulgar, se ben, na comunicación de carácter científico utilízanse tamén palabras deste vocabulario usual e non só do vocabulario científico (p. 39). En palabras de Galisson e Coste (1976):

A lingua usual é o conxunto dos medios lingüísticos habitualmente utilizados en situacións de comunicación (orais e escritas) da vida cotiá... Oponse á noción de norma porque deriva dunha constatación (o que se di), non dunha prescrición (o que se debe dicir) e ás nocións de lingua científica, linguaxe técnica, linguaxe profesional, linguaxe especializada que remiten a unha actividade específica, por conseguinte a un grupo social particular (p. 583).

O termo *uso* refírese á lingua practicada pola maioría dos usuarios dun grupo lingüístico correspondente a un grupo social determinado, á lingua que normalmente se fala nese grupo lingüístico. Está relacionado estreitamente con outros conceptos como dispersión, frecuencia e norma. O concepto de *dispersión* ou *rango* está directamente relacionado coa cantidade de ámbitos en que aparece un vocábulo e coa súa estabilidade na frecuencia de aparición, co que estes dous conceptos deben terse en conta no estudo analítico da produción lingüística e no establecemento de conclusións sobre a mesma (López Morales, 1984, p. 60). O concepto de *norma*, enténdese como o conxunto de prescricións que regulan o uso máis adecuado da lingua. Non existe oposición entre uso e

norma, pero o uso non sempre respecta a norma e, co paso do tempo e a medida que a frecuencia se estende amplamente, pode chegar a cambiala (López Morales, p. 60). Isto é o que sucede con certas palabras ou usos lingüísticos en xeral non aceptados polas normas, pero que se xeneralizan e rematan impoñéndose e sendo aceptados. O índice de uso obtense de combina-la frecuencia coa dispersión ($uso = dispersión \times frecuencia$), de modo que se incrementa a medida en que a frecuencia e a dispersión se elevan. Por exemplo, se a frecuencia do vocábulo *casa* é de 81 e a dispersión entre os diferentes ámbitos do que se extrae é de 0,92, o índice de uso será de 74,52 ($81 \times 0,92$). Así, pode darse o caso de que dous vocábulos coa mesma frecuencia teñan un índice de uso distinto, se a dispersión é distinta. En ocasións emprégase como criterio estatístico fiable para a elección de vocábulos na investigación e en didáctica.

García Hoz (1999, inicialmente, 1952) extraeu un *Vocabulario común* a partir do *Vocabulario usual*. Este *Vocabulario común* está constituído por 2181 palabras que aparecen na lingua familiar, na da rúa, na oficial e na da cultura en xeral, de forma que son empregadas usualmente en calquera situación en que se poda encontrar unha persoa. Suxire que se debería utilizar como base para a ensinanza inicial da lectura e escritura, xa que ocupan tres cuartas partes da linguaxe escrita.

O mesmo autor publicou en 1977 un *Vocabulario común fundido*, elaborado a partir de 3102 palabras dos *Vocabularios común general* e *común científico*, que cobren o 85% das frecuencias do vocabulario usual e o 83% das do científico. O *Vocabulario común fundido* está composto por 978 palabras que a xuízo do autor son comúns en calquera situación de comunicación cotiá ou en calquera situación de comunicación científica e que cobren o 86% das frecuencias do *Vocabulario común usual* e o 89% das frecuencias do *Vocabulario común científico*. Cada palabra deste *Vocabulario común fundido* vai acompañada da súa frecuencia no *Vocabulario común usual* e *común científico*.

1.2.6 Vocabulario específico

Galisson e Coste (1976) definen este vocabulario como aquel que se utiliza en situacións de comunicación oral e escrita que implican a transmisión de información relevante nun campo de coñecemento determinado (p. 511).

En parte corresponderíase co que García Hoz (1976b) denomina *Vocabulario general de orientación científica*, aquel que se limita a un campo propio da ciencia, que representa un maior grao de esixencia na súa precisión nos significados en contraposición ó común e usual. Vai acompañado dunha esixencia de exactitude e precisión, características dos campos científicos, técnicos e profesionais dos que o vocabulario específico se nutre. En moitas ocasións non é máis que a especificación, en base a un acordo previo entre quen a impón, do vocabulario usual, adquirindo este acepcións diferentes. Establécese, ademais, unha oposición entre vocabulario específico e xenérico en función do grao de extensión que separa a un do outro, de forma que o xenérico implica unha maior extensión, o que significa que un vocábulo xenérico pode substituír a un específico, pero non á inversa.

1.2.7 Vocabulario oral e escrito

Cabe establecer a distinción entre vocabulario oral e escrito en tanto que a experiencia mostra que nas etapas iniciais da aprendizaxe da lingua escrita un escolar seguramente recoñece máis palabras oralmente que na súa forma escrita, aínda que co tempo chegan a igualarse. Unha situación semellante pode producirse con alumnado que presente dificultades importantes na decodificación da lingua escrita. Chall (1987) estimou que o alumnado de primeiro curso de primaria coñece en torno a 6000 palabras, pero distingue entre as que coñecen e usan oralmente e as que recoñecen na escrita. Sinala que non é ata cuarto curso aproximadamente cando recoñecen na escrita 3000 palabras das que coñecían o seu significado oralmente en primeiro. Nalgúns traballos incluso se afirma que a linguaxe utilizada nos libros de texto é 10 veces máis complexa que a que utilizan na

conversación graduados universitarios (Alabama State Department of Education, 2001). Do mesmo xeito (Carlo et al., 2004), os escolares inmigrantes que acceden ó sistema escolar en cursos intermedios de primaria probablemente adquiran oralmente a segunda lingua con certa rapidez, pero, á vez, permanezan desfasados respecto ós escolares nativos que tiveron contacto coa lingua oral do país receptor dende o nacemento e coa escrita dende o inicio da escolaridade, a menos que adquiran estratexias e habilidades para aprender rapidamente palabras que encontrarán na lectura e que non usarán con frecuencia na oral.

En relación con este feito, que na lingua escrita podan encontrarse palabras que non soen usarse na oral, Corson (1995, 1997) demostrou que o léxico inglés utilizado en contextos de conversación cotiás difire significativamente do utilizado en textos escritos, tanto pola frecuencia como pola orixe das palabras. Parece que as 1000 palabras máis frecuentes en inglés falado cobren o 94% das palabras habituais nunha conversación informal (Schonell, Meddleton e Shaw, 1956), mentres que estas 1000 palabras máis frecuentes cobren o 74% dun texto inglés escrito (Carrol, Davies e Richman, 1971). Isto pode deberse a que a lingua oral é máis espontánea que a escrita, a cal precisa de máis tempo. Cando se escribe pénsase máis o vocabulario a usar e o texto escrito permite efectuar revisións, co que o habitual é que se empreguen palabras menos familiares, á vez que require a utilización dun maior número de palabras para transmiti-la mesma idea, porque aspectos como o ton non teñen cabida e outros non se poden dar por supostos, como o coñecemento de cuestións concretas sobre tema que se trata. Polo contrario, no intercambio oral pódense ir resolvendo este tipo de cuestións a medida que se desenvolve o devandito intercambio.

O vocabulario escrito pódese relacionar co concepto de vocabulario visual, constituído polo almacén de palabras que se recoñecen e entenden automaticamente sen necesidade de analizalas. É un almacén de léxico especializado no recoñecemento das representacións ortográficas das palabras que se leron con anterioridade. É un almacén de memoria a longo

prazo das palabras que se viron un número suficiente de veces como para memorizalas. Cada palabra posúe un “limiar de activación” diferente, que depende dunha combinación de factores como o número de veces que se encontrara escrita, que se verá afectado pola frecuencia, o tempo transcorrido dende a última vez que foi vista, etc.

Este tipo de vocabulario pode desenvolverse mediante experiencias de lectura repetidas, pero os procedementos utilizados inclúen máis concretamente a aprendizaxe de palabras illadas, de palabras en textos ilustrados ou mesturando os dous procedementos (presentación de palabras illadas e logo en contexto). Ó respecto, Stuart, Masterton e Dixon (2000) encontraron que en escolares de 5 anos a habilidade para segmentación fonolóxica, o coñecemento do alfabeto e das correspondencias, facilitaba a aprendizaxe do vocabulario visual. Por outra parte, os mesmos autores encontraron tamén que a aprendizaxe de palabras víase favorecida por un procedemento de ensinanza baseado en palabras illadas fundamentalmente.

O desenvolvemento dun amplo vocabulario visual móstrase como un importante factor que inflúe na aprendizaxe inicial da lectura, existindo experiencias de ensinanza de lectura nas que se parte da aprendizaxe visual de palabras, como é o caso dos denominados métodos globais, ou que se basean na aprendizaxe de palabras e frases, como é o caso de experiencias concretas de ensinanza de lectura precoz (Doman, 1991). A súa importancia tense demostrado igualmente na construción do significado de textos (Cooper, 1997, p. 168; Tan e Nicholson, 1997), na medida en que libera ó lector da análise da maioría das palabras con que se encontra ó seren palabras que se recoñecen e entenden automaticamente.

Como síntese, pódese establecer unha distinción entre vocabulario oral e escrito, o cal inicialmente e menor cuantitativamente que o oral, pero progresivamente vaise invertindo a relación. O vocabulario escrito pode relacionarse co vocabulario visual, composto polas palabras que se entenden automaticamente. Este vocabulario ten unha gran

importancia tanto na aprendizaxe da lingua escrita coma no desenvolvemento da comprensión lectora.

Despois do exposto ó longo deste apartado, non parece posible responder dun xeito simplista á pregunta que figura no encabezado (“¿Que significa coñecer unha palabra?”), máxime tendo en conta o comentado no punto 1.1.1 acerca de qué é unha palabra, xa que como se puido apreciar, tampouco é un concepto que resulte sinxelo definir. Parece que, como noutros ámbitos, será necesario acepta-la complexidade e ambigüidade destes conceptos e, á hora de valorar se unha persoa coñece unha palabra, seguramente sexa adecuado adoptar un enfoque que dalgún modo permita detectar ese coñecemento nos niveis máis básicos.

1.3 Relevancia do vocabulario

A pesar do status lingüístico ambiguo da unidade palabra e da dificultade para acoutar que significa coñecer unha palabra, nesta sección verase que ó vocabulario concédeselle unha importancia enorme tanto con carácter xeral coma no ámbito escolar.

Hai estudos que proporcionan evidencia de que o tamaño do vocabulario aparece como un factor significativo na predición da competencia xeral da linguaxe, ten un importante peso en tódalas habilidades lingüísticas, de forma que persoas con amplo vocabulario é probable que actúen mellor en tódalas outras tarefas lingüísticas que os que posúen un menor vocabulario (Meara, 1987; Meara e Jones, 1988; Zimmerman, Broker, Shaughnessy e Underwood, 1977). Neste sentido, por poñer un exemplo, Spring e Prager (1992), nunha investigación na que trataron de comprobar en que medida un programa de instrución incidía en identificación de ideas principais, empregaron un test de vocabulario para determinar se os grupos cos que traballaban eran comparables na súa habilidade verbal xeral, é dicir, utilizaron o dominio do vocabulario como indicador xeral da competencia

lingüística. O test constaba de 50 ítems de elección múltiple e, como resposta, en cada ítem os suxeitos tiñan que elixir entre cinco palabras o sinónimo da palabra estímulo.

No ámbito escolar é habitual que os educadores consideren o vocabulario como unha característica clave na linguaxe académica (Solomon e Rhodes, 1996). A investigación parece avalar esta impresión e ten demostrado que o tamaño do vocabulario correlaciona co éxito en tódalas áreas do currículo (Álvarez e Díez-Itza, 2000; Maning, 1999). Para Cummins (2002) probablemente sexa a manifestación máis clara ou o núcleo dunha dimensión xeral do dominio da lingua académica, relacionada coas capacidades cognitivas e que inflúe conxuntamente coas mesmas no desenvolvemento académico. Penetra tódalas áreas do currículo e a súa avaliación, tanto no referido ó vocabulario xeral coma ó específico das áreas, pode servir como un dos indicadores de en que medida se domina o contido desas áreas. A comprensión de conceptos académicos e o coñecemento léxico pódense considerar como dúas caras da mesma moeda e ámbolos dous van desde o simple recoñecemento ata a capacidade de explicalos, definilos, utilízalos e aplícalos en diversos contextos. Como ilustración desto, August e Hakuta (1997) citan a descrición da comunicación matemática no *New Standards* (1995), documento que recolle os niveis esixidos na escola elemental nas diferentes materias. Descríbena do seguinte xeito:

Usar termos, vocabulario e linguaxe matemáticos apropiados, baseándose no traballo conceptual previo: mostrar ideas de diversos modos, incluíndo palabras, números, símbolos, debuxos, cadros, gráficos, táboas, diagramas e modelos; explicar con claridade e dun modo lóxico as solucións ós problemas e apoiar estas solucións con probas, tanto de forma oral coma escrita; considera-la finalidade e o público ó comunicar; e comprende-las matemáticas ó le-las tarefas e outras fontes (pp. 120-121).

Podería dicirse que o vocabulario "(...) non constitúe toda a linguaxe, pero pode considerarse como o elemento material, e, por elo, constitúe a base da linguaxe" (García Hoz, 1977, p. 29), como tamén é o fundamento de toda tarefa escolar. Como sinalan Bruning, Schraw, Norby e Ronning (2005):

Entende-las palabras e saber como usa-las – coñecemento do vocabulario - , é un importante índice de dominio e coñecemento xeral. O coñecemento do vocabulario tamén inflúe na eficacia da aprendizaxe; un amplo vocabulario axuda ó procesamento cognitivo de diversas formas, mellorando a velocidade e comprensión lectora, e facendo máis precisa a expresión das ideas ó escribir ou falar (p. 314).

A modo de exemplo, cabe mencionar algunhas investigacións que relacionan o vocabulario co rendemento académico. Beck, Black e Doles (1985), analizaron a relación entre o Test de Vocabulario en Imaxes de Peabody e probas de rendemento con alumnado entre 6,3 e 11,5 anos, obtendo correlacións significativas entre 0,57 e 0,9, polo que conclúen que este test pode ser de utilidade como instrumento na identificación de dificultades no rendemento académico. Na mesma liña, Shaughnessy e Evans (1986) conclúen que o vocabulario medido co test Peabody é un instrumento útil para predici-lo rendemento académico de alumnado de bacharelato e de primeiro ano de facultade. Non obstante, Tillinghast, Morrow e Uhlig (1984), con alumnado de 4.º, 5.º e 6.º graos obteñen correlacións entre o vocabulario medido co test Peabody e o rendemento académico entre -0,18 e 0,67, polo que conclúen que o test Peabody é escasamente eficaz para predici-lo rendemento académico se se utiliza como único instrumento.

En España, Molina e García (1984, pp. 238-241) analizaron a relación entre diferentes variables e o rendemento académico con alumnado dos tres ciclos de EXB, encontrando que a variable que máis peso tiña nos tres ciclos era o vocabulario, avaliado mediante a proba do

mesmo nome do test WISC, xa que era a que mellor discriminaba entre grupos segundo o rendemento nos tres ciclos, a que correlacionaba máis alto (0,41, 0,59 e 0,59, correlativamente nos tres ciclos) e a que máis contribuía á varianza do criterio (24,07%, 34,5% e 34,68%, respectivamente).

1.3.1 Correlación coa intelixencia medida polos test

Abordouse esta cuestión sucintamente na introdución a esta primeira parte da tese, sinalando que o vocabulario é considerado coma a manifestación máis directa do coñecemento do mundo, á vez que é unha variable cun importante peso nos instrumentos destinados á medición da intelixencia xeral. Así, ademais de que gran parte das escalas de intelixencia inclúen unha subproba de vocabulario (WISC, MSCA, etc.), existen diversidade de traballos que aportan evidencia no sentido de que o vocabulario medido con probas como o Test de Vocabulario en Imaxes de Peabody, Test de Denominación de Boston, Fluencia do Weschler ou o Test de Conceptos Básicos de Boehm, correlaciona coa capacidade intelectual, segundo é avaliada polas probas de intelixencia ó uso. Deseguido coméntanse brevemente algunhas investigacións que proporcionan información sobre este tema, aínda que en todas as que se referencian utilizouse o Test de Vocabulario en Imaxes de Peabody para a valoración do vocabulario e, por conseguinte, é necesario ter en conta .que o vocabulario que se mide é esencialmente pasivo.

Obtivéronse correlacións positivas e significativas con alumnado de Educación Infantil entre as puntuacións do test Peabody e as da Escala Stanford-Binet (Alpeter, 1985; Carvajal, Hardy, Harmon e Sellers, 1987), así como entre as puntuacións do test Peabody e do test WPPSI (Vance, West e Kutsick, 1989), concluíndo os autores neste último caso que as puntuacións no Peabody poden ser utilizadas para predecir as puntuacións no WPPSI. Tamén se obtiveron correlacións significativas con escolares de 6 e 16 anos entre as puntuacións do test Peabody e do WISC-R (Alpeter, 1989; Halpin, Simpson e Martin, 1990), concluíndo os

autores que o test Peabody é un instrumento útil para predici-lo rendemento no WISC-R; obtivéronse igualmente correlacións significativas entre as puntuacións do test Peabody e as da Escala Stanford-Binet (Carvajal, McVey, Sellers e Weyand, 1987). Do mesmo modo, autores como Carvajal, Shaffer e Weaver (1989) obtiveron correlacións altas, de arredor de 0,80 con suxeitos de 24 a 36 anos entre a puntuación no test Peabody e a puntuación total no WAIS-R, co que o consideraron un instrumento válido para efectuar unha valoración da capacidade intelectual.

Non obstante, esta relación non sempre aparece de forma clara e diferentes investigacións póñena en dúbida ou, cando menos, matízana. Así, Alpeter (1985) comprobou con escolares de 2 anos e 6 meses a 5 anos e 11 meses que as puntuacións medias no test Peabody subestimaban as puntuacións medias da Escala Stanford-Binet e só un 55% dos casos eran correctamente clasificados en base ás puntuacións obtidas no test Peabody, o que a xuízo do autor suxire que este instrumento pode ser utilizado na valoración do funcionamento intelectual en educación infantil, pero con cautela. Tamén se comprobou que era cuestionable a eficacia do test Peabody para detectar suxeitos intelctualmente sobredotados en idades inferiores a 7 anos, polo menos, se se utilizaban como criterio instrumentos convencionais para medi-la intelixencia, como a Escala Stanford-Binet (Hayes e Martin, 1986). Algo semellante ocorría con escolares entre 2 e 15 anos, xa que, se ben as puntuacións no test Peabody e na Escala Stanford-Binet correlacionaban significativamente, producíanse erros na clasificación do rendemento intelectual nun 50% dos casos se para a clasificación se utilizaba o test Peabody (Tarnowski e Nelly, 1987). Sucede o mesmo con escolares entre 6 e 11,11 anos: as correlacións entre as puntuacións conseguidas no test Peabody e no WISC-R foron significativas (0,78; 0,79; 0,68 nas tres escalas), pero o test Peabody subestimaba as puntuacións no WISC-R dos suxeitos que se situaban na zona máis baixa do rango, mentres que sobreestimaba as dos que se situaban na parte alta do rango (Alpeter e Handal, 1986). Con mozos e adultos de 18 a 48 anos as correlacións entre o

rendemento no test Peabody e no WAIS-R foron igualmente significativas, pero só en menos da metade dos casos eran clasificados correctamente mediante o test Peabody en canto ó funcionamento intelectual, xa que se sobreestimaban as puntuacións do WAIS-R nos casos que obtiñan puntuacións máis baixas e subestimaban a dos casos con puntuacións máis altas (Alpeter e Johnson, 1989).

Polo tanto, parece que existe unha correlación entre estes dous construtos, pero é necesario entendela con precaución porque non en tódolos casos se obtivo evidencia de que, cando menos, co instrumento utilizado máis habitualmente para a valoración do vocabulario (test Peabody) fose posible predici-lo rendemento intelectual a partir do nivel de coñecemento do vocabulario, tal como se entende que o miden os tests de intelixencia.

1.3.2 Lingua oral

Inclúese este punto co obxecto de resalta-la importancia do vocabulario en relación coa lingua oral nas súas vertentes comprensiva e expresiva. Pero, con tódalas diferenzas existentes entre a lingua oral e a escrita, pódese considerar a importancia da palabra nos mesmos termos en ámbolos dous planos. Por elo, cabe aduci-lo mesmo tipo de argumentos que se exporán con certa amplitude nos puntos 1.3.3, 1.3.4 e 1.3.5 en relación coa lingua escrita. Aquí unicamente se fará referencia a grandes trazos a esta cuestión.

Intuitivamente parece obvio que na medida en que unha persoa sexa capaz de entender e utilizar un maior número de palabras, maiores van a ser as súas posibilidades para entender-lo discurso elaborado por outras persoas, así como de expresarse con maior precisión, calidade, etc. Isto, á súa vez e xunto co desenvolvemento das habilidades en lingua escrita, posiblemente repercuta nas súas posibilidades de ampliar coñecementos, de desenvolver unha maior capacidade de elaboración conceptual e de conseguir mellora-las súas habilidades intelectuais.

Retomando o exposto no apartado 1.1, a unidade lingüística palabra, como mostra da lingua oral, pódese considerar como un instrumento por medio do que os humanos construímo-la realidade e por medio do que lle damos sentido ó mundo que nos rodea, o que explica que, por exemplo, nalgunhas tribos do trópico non dispoñan do vocábulo *deserto* ou que entre os esquimós dispoñan de catro vocábulos para denomina-la *neve* (neve no chan, neve caendo, neve nunha cifra e neve nun montón). Tamén permite “manipular” mentalmente o entorno sen necesidade de estar en contacto con el. Na argumentación de Vygotsky, a palabra reflicte o pensamento, coroa a acción posta en marcha a partir dunha motivación, primeiramente en forma oral e posteriormente en forma escrita (1962, p. 153).

Como se explicará no punto 2.1.2 sobre o almacenamento do léxico mental, na vertente comprensiva, cando unha persoa escoita unha palabra realiza unha análise fonolóxica que lle permite identificar automaticamente os segmentos de que consta; realizada esta análise, combina o resultado coas representacións almacenadas no léxico mental o que permite o recoñecemento da secuencia de fonemas; recoñecida esta secuencia, faise posible a representación morfolóxica e accédese ó significado e ó listado de termos asociados á palabra concreta. No que respecta á vertente expresiva, unha vez formada unha idea do que se quere dicir, para o que é necesario dispoñer de coñecementos sobre o mundo en xeral, entre os que se encontran os de vocabulario, ponse en marcha a codificación gramatical, con asignación de roles temáticos (axente, paciente, lugar, etc.) e sintácticos (suxeito, predicado, etc.), codificación que precisa da recuperación léxica para a construción das oracións que se van a emitir. Esta recuperación léxica consiste en facer dispoñibles a forma das palabras, o seu significado, a súa función sintáctica e a súa estrutura silábica. En lingua oral, coma sucede tamén na escrita, a utilización activa esixe un coñecemento máis profundo da palabra, en tanto que hai que evoca-la forma completa, co que se ve limitada a expresión e o que comunmente se denomina “falar con propiedade” se o “caudal” léxico é insuficiente.

1.3.3 Aprendizaxe inicial da lingua escrita

Neste punto tratarase con maior profundidade un ámbito fundamental como é o da lingua escrita, expoñendo os resultados dalgunhas investigacións que parecen achegar evidencia acerca da relación entre o vocabulario e a aprendizaxe da mesma. En primeiro lugar, coméntanse tres investigacións españolas e a continuación algunhas do ámbito anglosaxón.

Escoriza (1986) estudou o poder dunha serie de variables para predici-lo rendemento en lectura en 1.º de EXB. As probas seleccionadas miden factores perceptivos, lingüísticos e conceptuais. Formulou catro hipóteses e analizou os datos que obtivo mediante os procedementos seguintes: matriz de correlacións, análise discriminante e regresión múltiple paso a paso. Os resultados indican que é máis efectivo utilizar unha batería de probas, pero as medidas que presentan un maior poder para predici-lo rendemento en lectura entre as utilizadas na investigación son, por un lado, o subtest de vocabulario do test WISC e, por outro lado, unha tarefa empregada para comproba-lo coñecemento de letras, consistente en busca-la letra igual á ofrecida como modelo e que o autor considera como o tipo de resposta de menor dificultade nunha escala (identificar dúas letras como iguais ou diferentes, nomear letras, copiar letras).

Jiménez e Artiles (1989, 1990) aplicáronlle a 60 alumnos ó inicio de 1.º de EXB a Escala McCarthy, o Test Boehm de Conceptos Básicos, a Proba de Percepción Visual de Frostig, a de Reproducción de Estructuras Rítmicas de Stambak, a proba de Brunet-Lezine sobre o Coñecemento do Propio Corpo, o Test de Dominancia Lateral de Harris, o Test Motor de Ozoreski e a Proba de Discriminación Fonemática de Rodríguez Jorin. Ó inicio de 2.º de EXB, aplicáronse probas de lectura (velocidade, exactitude e comprensión), ditado de palabras e frases e composición escrita. Os resultados das probas de inicio de 1.º sometéronse a análise estatística, mediante a técnica de compoñentes principais, extraendo a estrutura interna dos factores lingüístico, estruturación visomotriz, organización cognitivo-

conceptual e memoria reprodutiva. O poder de cada un dos factores para predicir diferentes variables de lingua escrita valorouse mediante unha análise de regresión múltiple, resultando se-los factores con maior influencia en lectura e escritura o factor lingüístico e o factor de organización conceptual. A teor dos datos obtidos nesta investigación, parece que son os factores nos que se pode considerar que o vocabulario ten un peso maior os que predin en maior medida a aprendizaxe da lingua escrita, aínda que estes factores estean conformados por máis variables que as propiamente relacionadas co vocabulario.

García Vidal (1999) realizou un estudo con 829 escolares de Andalucía que cursaban primeiro de ensinanza primaria, no que avaliou o poder dunha serie de variables para predici-lo “descifrado”, a “lectura de palabras” e a “comprensión de palabras e frases”, atendendo ó método usado para a ensinanza inicial da lectura, á titularidade do centro escolar e ó sexo. En términos xerais, as variables con maior peso na predición foron o coñecemento de letras e números. Non obstante, había un conxunto de variables que tamén interviñan de forma importante nesa predición. Dentro deste conxunto de variables encontrábase ó coñecemento léxico, avaliado mediante tarefas de comprensión e expresión verbal, sinónimos, antónimos e categorizacións verbais. De forma resumida, a variable coñecemento do léxico aparecía como unha variable que contribuía á predición, en primeiro lugar, da variable criterio “descifrado”, no caso dos suxeitos que se iniciaban na lectura tanto mediante métodos sintéticos como analíticos, no grupo de nenos e no alumnado de colexios públicos. En segundo lugar, tamén era unha variable que participaba na predición da “lectura de palabras” na mostra xeral e na de colexios concertados. En terceiro lugar, participaba na predición da variable “comprensión de palabras e frases” no grupo de nenos, no grupo de centros públicos e no grupo de centros concertados.

No ámbito anglosaxón, os datos parecen ir na mesma dirección. Os resultados obtidos por Pikulski e Tobin (1989) nun estudo de carácter lonxitudinal con alumnado de 4 anos aproximadamente e alumnado de 6.º grao, Scarborough (1989), nun estudo con

alumnado de 2.º grao e Stanovich, Nathan, e Vala-Rossi (1986), con alumnado de 3.º e 5.º graos, parecen indicar igualmente que unha das variables que mellor predin a habilidade lectora é o vocabulario (xunto con antecedentes de problemas de lectura na familia, a conciencia fonolóxica e habilidades relacionadas coa alfabetización temperán ós cinco anos). Con todo, parece que o vocabulario na súa vertente expresiva resulta ser máis útil na predición da habilidade lectora que na súa vertente receptiva (Scarborough, 1991). A esta conclusión de que o vocabulario expresivo é unha variable máis potente para a predición que o receptivo chegaron tamén Snow, Burns e Griffin (1998) despois de revisar cinco estudos nos que o vocabulario expresivo se valorou mediante unha tarefa de denominación (correlación de 0,45) e o receptivo mediante a tarefa que consistía en indica-lo debuxo que se solicitaba (correlación de 0,36).

Das investigacións revisadas cabe concluír que o vocabulario, valorado mediante diferentes procedementos, é unha das variables que inciden na aprendizaxe inicial da lectura, sen chegar a se-la máis importante en tódolos casos. Dos datos que se manexaron non é posible extraer co mesmo grao de seguridade a conclusión de que o vocabulario sexa unha variable con tanta importancia na aprendizaxe inicial da escritura, porque unicamente no estudo de Jiménez e Artiles se aborda a relación con esta aprendizaxe.

1.3.4 Comprensión lectora

Existe un amplo consenso acerca de que para comprender un texto é necesario construí-lo modelo da situación á que se refire o texto ou modelo referencial (Díaz e De Vega, 2003, p. 144). Prodúcese o que se pode considerar un intercambio entre o contido do texto e a persoa que le, de forma que esta achega o seu coñecemento para poder comprender. En palabras de Eco (1994), "Todo texto é unha máquina preguiceira que lle pide ó lector que lle faga parte do seu traballo. Pobre do texto se dixera todo o que o seu destinatario debiera aprender: non acabaría nunca" (p. 11). Comprender un texto implica

construír relacións simples entre ideas (proposicións) con palabras do texto para construí-la microestrutura, xerarquiza-las ideas para construír un significado global do texto (macroestrutura) e recoñece-la trama de relacións entre ideas globais que dan lugar a diferentes tipos de organización textuais (superestrutura). Finalmente, como se sinalou ó inicio deste parágrafo, é necesario construí-lo modelo referencial de situación. Ruddell (1994) propón a seguinte definición:

Defino a comprensión, entón, como un proceso no que o lector constrúe o significado mentres interactúa co texto, ou despois, mediante a combinación de coñecemento previo e experiencia previa, información do texto, a postura que o lector adopta en relación co texto e as interaccións sociais e comunicación social inmediatas, recordadas ou anticipadas (p. 415).

As variables que inciden na comprensión lectora son diversas. Entre elas, autores como Alonso e Mateos (1985), Baddeley, Logie e Nimmo (1985), Cooper (1997), García Madruga, Martín, Luque e Santamaría (1995), Sánchez (1993, 1998), Snow, Burns e Griffin (1998) ou Valle (1991) citan os coñecementos previos de distinta índole, resultado da experiencia acumulada nos diferentes ámbitos da vida (desenvolvemento conceptual e de vocabulario, coñecemento do mundo en xeral que facilita a resolución de ambigüidades sintácticas/coñecemento sintáctico, coñecemento das estruturas textuais, coñecemento en forma de esquemas sobre a acción humana e sobre o mundo físico, coñecementos específicos sobre un tema concreto, etc.), pero tamén a capacidade para activalos ó ler un texto, estratexias metacognitivas de autorregulación da comprensión, memoria de traballo, procesos xerais de acceso ó léxico, etc.

Respecto do coñecemento previo, Tierny e Cunningham (1984), despois de realizar unha extensa revisión de investigacións sobre intervención en comprensión lectora, insisten

na existencia dunha relación causal entre coñecemento previo e comprensión lectora, entendido o coñecemento previo, en termos xerais, como o conxunto de contidos aprendidos ó longo da súa vida por unha persoa, ou, a xuízo de Harris e Hodges (1995), como o coñecemento que procede da experiencia acumulada por un individuo.

Máis concretamente, en investigacións como as de Spillich, Vesonder, Chiesi e Voss (1979) e Post, Greene e Bruder (1982, citados en Gutiérrez-Calvo, 1999), comprobouse que o rendemento dos suxeitos despois de ler ou oír un texto sobre béisbol variaba en función do coñecemento previo sobre o tema. O rendemento en tarefas de resume oral, resume escrito e resposta a preguntas foi peor no caso dos suxeitos que dispoñían dun menor coñecemento que o rendemento dos que tiñan un maior coñecemento. Foi peor igualmente o rendemento en canto á calidade dos xuízos e inferencias que se referían a cuestións específicas relacionadas co tema, o que non ocorría no caso de informacións complementarias que aparecían no texto, pero que non se referían a cuestións específicas do deporte descrito, como, por exemplo, as condicións atmosféricas. Na mesma liña, Stahl (1991), analizou o efecto do coñecemento previo e do vocabulario, tanto xeral coma específico no tema, sobre o recordo de aspectos do contido dun artigo dunha revista sobre béisbol con alumnado de aproximadamente quince anos (10.º grao, equivalente a cuarto de ESO). Encontrou que o coñecemento do tema e do vocabulario afectaban de forma independente á comprensión (o coñecemento do tema sobre o tipo de información recordada e o vocabulario sobre a cantidade de contido recordado).

Tamén Peterson (1993) investigou a relación entre coñecemento previo, comprensión lectora, decodificación e comprensión oral. Para elo, avaliou o coñecemento sobre béisbol de 135 persoas (vocabulario e procedementos). A continuación, unha vez lido un texto sobre o tema, avaliou a comprensión lectora (preguntas literais e inferenciais), a comprensión oral e a decodificación (ler en voz alta unha lista de 100 palabras extraídas do texto lido). Así mesmo, levou a cabo outro estudo seguindo o mesmo procedemento, pero utilizando un

texto sobre ordenadores. En base ós datos obtidos, parece claro que o coñecemento previo correlacionaba máis positivamente coa comprensión lectora e oral que coa decodificación e, por outra parte, os resultados parecen suxerir que a comprensión variaba en función do tema, probablemente debido ó coñecemento previo sobre o mesmo. Chegouse á mesma conclusión xeral sobre o papel dos coñecementos previos na comprensión lectora noutros múltiples traballos, entre os que cabe citar os de Gutiérrez-Calvo e Castillo (1997), Gutiérrez-Calvo, Eysenck e Castillo (1997) e Gutiérrez-Calvo, Eysenck e Estévez (1994).

Xa se comentou que na comprensión lectora interveñen diferentes variables e que o coñecemento previo é unha destas variables. O coñecemento previo, á súa vez, constrúese e exprésase por medio de diferentes linguaxes, principalmente a verbal, da que o vocabulario é un dos compoñentes, pero, ademais, outros tipos de coñecemento como os esquemas, por exemplo, teñen forma de proposicións que inclúen palabras. Se a lectura é un proceso mediante o que se interpreta un referente, cando se lle asigna a un estímulo significativo un significado polo que os símbolos lidos pasan a ser signos, implica que o lector necesita un dominio mínimo do léxico referido ó tema do texto, senón a lectura convértese nun mero descifrado. Ó respecto, autores como Carver (1994) sinalan que na lectura dun texto con máis dun 2% de palabras descoñecidas bloquéase a comprensión, así como a aprendizaxe de novas palabras. Freebody e Anderson (1983) encontraron que con alumnado de sexto de primaria esta porcentaxe era de ata o 15%, mentres que en quinto situábase entre o 3% e o 6%. De tódolos modos, pódese concluír que a proporción tolerable de palabras descoñecidas en relación coa dificultade para comprender un texto depende do rol destas palabras no texto e do propósito da lectura, segundo pon de manifesto Nagy (1988).

Existe un número considerable de investigacións que suxiren que o coñecemento do vocabulario é un dos principais determinantes da comprensión lectora con grupos étnicos distintos (Anderson e Freebody, 1979; Carlo et al., 2004; Hall e Tirre, 1979; Raphael, Freebody, Fritz, Mayers e Tirre, 1980), con escolares de ensinanza primaria (Alonso, Carriedo

e González, 1992; Callahan e Drum, 1984; Carlo et al., 2004; Nacional Institute of Child Health and Human Development, 2000), de ensino secundario (Alonso, Carriedo e González, 1992; Espin e Foegen, 1996; O'Brien, 1986), universitarios (Farley e Elmore, 1992), adultos (Baddeley, Logie e Nimmo-Smith, 1985) e con aprendices dunha segunda lingua (Carlo et al., 2004; Nagy, 1997; Quian, 1999; Read, 1997; Verhoeven, 1990).

Segundo Johnston (1989), nos estudos sobre a capacidade lectora, as dificultades en vocabulario explican o 80% da varianza. Anderson e Freebody (1979) propuxeron tres hipóteses explicativas destes resultados. A instrumentalista sostén que o coñecemento das palabras dun texto permite a comprensión do mesmo, mentres que o descoñecemento das devanditas palabras impide a lectura adecuada. A hipótese da aptitude, mantén que coñecerlo vocabulario é un indicio máis do CI, que sería o auténtico factor explicativo da comprensión. A hipótese do coñecemento previo mantén que o coñecemento do vocabulario é un indicio da importancia das estruturas conceptuais previas (esquemas). Ó respecto Johnston (1989) inclínase pola terceira, xa que canto maior sexa o coñecemento previo do lector, maior será a probabilidade de que coñeza as palabras relevantes, de que realice inferencias adecuadas mentres le e de que constrúa modelos de significado correctos. Se o vocabulario é indicativo do coñecemento previo, entón un vocabulario limitado indicaría unha capacidade limitada para realizar inferencias correctas e construí-los modelos de situación que posibilitan unha lectura de arriba-abaixo eficaz. A compoñente de razoamento na lectura veríase restrinxida ante a falta de datos para resolver os problemas que supón o texto.

Cabe así mesmo explica-las diferenzas en canto a comprensión lectora en función do grao de coñecemento do vocabulario porque este coñecemento, xunto coa habilidade de análise visual e de conversión de grafemas en fonemas, permite recoñecer con rapidez e precisión as palabras o que permite o acceso ó significado das mesmas. Cando unha persoa precisa adica-la maior parte da súa capacidade cognitiva a identifica-las palabras

descoñecidas, encontrará enormes dificultades para construír unha representación adecuada do mundo ó que se refire o texto, se non imposibilitan este proceso.

En España, Espín (1986, 1987) estudou a relación entre doce variables e a comprensión lectora, utilizando unha mostra de escolares de 2.º de EXB pertencentes ás clases sociais media-alta e baixa (402 segundo a publicación de 1986 e 437 segundo a de 1987). Das 12 variables, sete estaban relacionadas coa expresión oral e as cinco restantes eran a actitude ante a lectura, a intelixencia (medida co test Raven Color), o vocabulario oral (valorado coa proba de Vocabulario do WISC), o vocabulario escrito (medido cunha tarefa de busca-la palabra contraria) e comprensión oral. Estas variables considerounas como variables que podían predici-la comprensión lectora. Como técnica de análise estatística utilizou a regresión múltiple, medindo o peso das cinco variables mencionadas non relacionadas coa expresión oral. Coa mostra total as variables que máis contribuían foron o vocabulario escrito (24,44%), vocabulario oral (6,54%) e intelixencia (3,84%). No caso da mostra de clase media-alta era novamente o vocabulario escrito a variable que explicaba unha maior porcentaxe da varianza do criterio (35,53%), seguida da actitude en lectura (4,42%), mentres que na clase baixa as que máis contribúen a explica-lo criterio son o vocabulario escrito (25,43%) e vocabulario oral (4,35%). Incluindo na análise tódalas variables con peso na predición, era igualmente o vocabulario escrito a variable que mellor predicía o criterio (35,53% da varianza explicada na mostra global, 24,42% na clase media-alta e 25,43% na clase baixa). O vocabulario oral resultou se-la segunda variable con maior peso tanto na mostra global (6,16%) como na mostra de clase baixa (4,42%). Eliminando da análise o vocabulario escrito, foi o vocabulario oral a variable con maior poder de predición (18,05% da varianza explicada na mostra total, 7,08% na mostra de clase media-alta e 11,03% na mostra de clase baixa) (pp. 83-94).

Schoonen, Hulstijn e Bossers (1998) analizaron a contribución á comprensión lectora en L1 e L2 do coñecemento do vocabulario e o coñecemento de estratexias metacognitivas

con escolares holandeses de 6.º (12 anos), 8.º (14 anos) e 10.º (16 anos) graos. No caso do alumnado de 6.º, en L1 o vocabulario explica o 60% da varianza de comprensión lectora, medida cunha proba de elección múltiple, mentres que en 8.º é o 45% e en 10.º o 48%.

Tamén se comprobou que a instrución en vocabulario incrementa a comprensión lectora (Carlo et al., 2004), a habilidade para establecer conexións causais (Medo e Ryder, 1993) ou para a identificación de ideas principais (Nagy, 1988; Stahl e Fairbanks, 1986), á vez que o incremento en comprensión lectora se xeneralizaba a textos diferentes do utilizado inicialmente (Stahl e Fairbanks, 1986).

Parece, pois, que está demostrada a importancia do vocabulario na comprensión lectora. Non obstante, debe terse en conta que escolares co mesmo nivel de vocabulario poden dispoñer de diferentes graos de comprensión, entrando aquí en consideración o resto das variables que inciden na mesma. Así, determinadas persoas poden fracasar con textos de vocabulario familiar, como constatou Sánchez (1993) ó utilizar textos con vocabulario coñecido por alumnado de 6.º e 7.º de EXB, xa que se producen importantes diferenzas en canto a comprensión entre estes escolares.

Por outro lado, como se explicará nos puntos 2.1.1 e 2.1.4 do capítulo 2, gran parte do vocabulario do alumnado medio procede da lectura, a partir da inferencia do significado de expresións novas polo contexto, de modo que o incremento da lectura pódese considerar como o medio máis eficaz para eleva-lo grao de coñecemento do vocabulario (Anderson e Nagy, 1993; Nagy, 1988; Nagy, 1997; Stenberg, 1987, entre outros). Pero, para que se produza esta influencia, son necesarias unha serie de condicións, que se poderían compendiar en coñecementos lingüísticos, coñecementos en xeral sobre o mundo e coñecemento de estratexias para utiliza-lo contexto. Neste sentido, as relacións entre comprensión e vocabulario móstranse moi complexas: cun limitado vocabulario é difícil a comprensión, pero, á súa vez, a lectura é a fonte principal do vocabulario, para o que é preciso que sexa comprensiva.

A pesar de tódalas matizacións, a aprendizaxe de novos conceptos é esencial para o desenvolvemento da comprensión. A habilidade para inferir e reter novas palabras en xeral, depende en parte do coñecemento previo doutras palabras e conceptos (Stanovich, 1986). Incluso en idades moi baixas, a habilidade para entender e recordar significados de novas palabras, depende en gran medida do desenvolvemento do vocabulario dispoñible (Robbins e Ehri, 1994), aínda que sexa preciso considerar outras variables como son o coñecemento das estruturas textuais, as diferenzas cualitativas no coñecemento previo sobre o contido, etc.

Como conclusión, poderíase dicir que para comprender un texto, inicialmente é necesario recoñecer con rapidez e precisión as palabras escritas, o que supón coñecementos sobre vocabulario. Pero, un nivel de comprensión máis profundo implica, non só coleccionar significados de palabras e establecer relacións entre elas para posibilita-la elaboración de ideas, senón a posibilidade de establecer relacións entre o que o texto ofrece e o coñecemento previo. Parece que canto maior sexa o coñecemento do vocabulario relevante, maiores serán as posibilidades de establecer inferencias pertinentes e construír modelos de situación máis correctos. Pódese entender que o coñecemento das palabras actúa facilitando o establecemento de pontes en canto ó significado, de modo que se integren os coñecementos novos noutros xa adquiridos e se relacionen significados.

1.3.5 Composición escrita

En principio, dadas as relacións existentes entre a lectura e escritura (Cooper, 1997, pp. 335-336), que son dous procesos inextricablemente unidos (Figura 1.5.), en base ás mesmas evidencias aducidas no punto anterior (1.3.4), cabe pensar que o vocabulario sexa un factor con gran peso no desenvolvemento da calidade da expresión escrita. Poderíase dicir que incluso a habilidade para escribir con efectividade depende máis dun adecuado nivel de vocabulario que a lectura (Brynildssen, 2000), porque no proceso de produción dun

texto non é posible analiza-lo contexto no que é utilizada unha palabra (a persoa que escribe esta creando). O autor ou autora do texto, cando escribe, debe ser capaz de recordar espontaneamente palabras que coñece máis que dunha forma superficial; debe coñecelas o suficiente como para poder usalas correctamente. Neste sentido, a escritura depende da habilidade para evoca-las palabras, co que a amplitude do vocabulario terá unha influencia directa na precisión e en xeral na calidade do que se escribe. Un vocabulario rico permite que un escritor poda expresa-la riqueza do seu pensamento, pero esto non é resultado do uso de palabras especialmente cultas ou pouco habituais senón do uso das palabras máis acertadas.

Procesos	Lectura	Escritura
Planificación	Ter un propósito para ler Xerar coñecementos previos	Ter un propósito para escribir Xerar coñecementos previos
Composición	Lectura e construción do significado do texto	Escritura e composición do texto
Edición	Relectura, reflexión e revisión do significado do texto	Relectura, reflexión e revisión
Monitorización	Remate da construción do significado do texto	Remate da composición

Figura 1.5.: Comparación do procesos de lectura e escritura, segundo Cooper (1997)

Dalgún modo (Carter, 1986; Giammateo e Basualdo, 2003), sempre se recoñeceu a importancia do léxico no proceso de ensinanza e aprendizaxe da expresión escrita, pero, en xeral non semella que se lle dedicase a suficiente atención a este aspecto da linguaxe. Habitualmente non parece haber unha práctica sistematizada orientada á ampliación e enriquecemento do vocabulario nin para que os escolares aprendan a utilizalo como

instrumento facilitador da comprensión e produción de textos en xeral, como tampouco, máis en concreto, de textos disciplinares, fundamentalmente dos de tipo expositivo, que poden considerarse como os máis abundantes no currículo escolar. Como exemplo desta falta de atención, na *Writing Report Card for the Nation and States* (Greenwald, Persky, Campbell e Mazzeo, 1999) non se menciona o rol do vocabulario ou a selección de palabras. Non obstante, nun modelo articulado da escritura como é o de Flower e Hayes (1994) a selección de cada palabra cobra importancia nas tres fases do proceso (planificación, produción e revisión), de modo que se constitúe nun elemento que condiciona o que o autor desexa e pode expresar. De igual forma que un título condiciona o contido dun texto ou unha frase tópica condiciona as opcións dun parágrafo, cada palabra limita dalgún modo a selección das subseguinte. Poderíase dicir que o bo uso do vocabulario significa usa-la palabra apropiada no momento oportuno e pola razón idónea.

En relación co que se acaba de sinalar, caben varias precisións (Carter, 1986). En primeiro lugar, os diferentes tipos de palabras xogan papeis distintos no que atinxe á efectividade do texto: é máis probable que abunden palabras expresivas ou non nucleares (con valor connotativo) nun texto descritivo que noutro que conteña unha secuencia de acción da vida cotiá, no que é probable que se utilice un maior número de palabras nucleares ou non expresivas (non introducen matices, como ocorre con palabras como *mesa* ou *comida*). En segundo lugar, o número e o tipo das palabras gramaticais empregadas parece que contribúe á coherencia dun texto e é altamente relevante para a efectividade do mesmo, pero o seu mero relato (a mera análise dos elementos de cohesión) non explica por qué algúns textos son percibidos como mellor organizados que outros. En terceiro lugar, parece conveniente fomenta-la utilización na escritura de palabras abertas (substantivos, adxectivos, verbos e adverbios) de forma suficiente e variada, pero a relación entre *type* (número de palabras/formas diferentes nun texto) e *tokens* (número total de palabras nun texto) non resulta unha medida efectiva da expresividade, non asegura en si mesma a

coherencia e non aclara as razóns polas que un texto se pode considerar mellor ou peor no relativo á utilización do vocabulario. En cuarto lugar, a utilización das relacións de sentido entre palabras (hipónimos, sinónimos, antónimos ...) é tan importante para a efectividade do texto coma o concepto que expresa o significante.

Un elemento máis a ter en conta é que cando se escribe faise cun propósito, para unha audiencia e isto inflúe na selección das palabras. Por exemplo, os textos académicos precisan dunha linguaxe apropiada segundo a temática e a materia, distinta da escritura de ficción ou dunha descrición, pero incluso unha lista de compra pode esixir unha coidadosa selección se vai a ser utilizada por alguén que non pertenza ó círculo máis achegado de quen a escribe. En consonancia con esto, a xuízo de Giammateo e Basualdo (2003), a instrución en competencia léxica, xunto coa instrución sobre cómo servirse dese competencia como ferramenta cognitiva, pode promover unha mellora na comprensión e produción de textos disciplinares. A capacitación nos niveis léxico e textual beneficia non só á competencia lingüística senón que facilita o acceso ó coñecemento e favorece o desempeño curricular.

Deseguido comentaranse os resultados dos traballos de dous grupos de investigadores, co obxecto de ilustrar esta importancia do vocabulario na expresión escrita. Duin (1983) e Duin e Graves (1986, 1987, 1988) realizaron unha serie de estudos co fin de investiga-lo efecto da ensinanza do vocabulario na fase de planificación dun texto sobre o uso do vocabulario no propio texto escrito e na calidade do mesmo. Na investigación de 1983, Duin traballou con escolares de cuarto e sexto de ensino primario, ensinándolles 10 palabras relacionadas cun tema sobre o que terían que escribir. En termos xerais, a calidade das producións do grupo experimental incrementouse, pero, apareceron diferenzas en función do nivel de escolarización: o alumnado de cuarto utilizaba como media un 80% das novas palabras aprendidas, mentres que os de sexto facíano nun 40%, aínda que os de cuarto non sempre as utilizaban adecuadamente e os de sexto mostraban un uso máis apropiado. Na investigación de 1987 ensináronlle 13 palabras relacionadas co tema do

espazo a escolares dunha idade equivalente a 1.º de ESO. A un grupo instruírono durante seis días utilizando tarefas como preguntas, discusión, gravacións, lectura de pasaxes, utilización de memorandos, anotación de palabras, libros de rexistros, etc., combinando as actividades de vocabulario con escritura. Con outro grupo utilizaron o mesmo tipo de instrución de vocabulario, pero sen a tarefa de escritura. Con terceiro grupo soamente utilizaron a instrución en vocabulario con fichas. Os resultados mostraron que o rendemento do primeiro grupo foi superior, seguido polo do segundo grupo, co que os autores concluíron que a ensinanza dun grupo de palabras relacionadas antes de iniciar a escritura dun texto no que as palabras poden ser utilizadas, favorece unha mellora na calidade dos ensaios.

Giammateo e Basualdo (2003) levaron a cabo unha investigación con alumnado de entre 16 e 17 anos na que partían da hipótese de que un mellor desempeño léxico e comunicativo favorece a produción escrita, contribúe a unha maior capacidade argumentativa e estimula o desenvolvemento do pensamento crítico. Trataron de indaga-lo uso do vocabulario abstracto, pero común a diferentes áreas, que é utilizado en operacións como comparar, fundamentar, argumentar, etc., indispensables para a comprensión e produción de textos propios da actividade académica. Valoraron o desempeño léxico, o recoñecemento e utilización de lexemas illados e en contexto, mediante probas de emparellar, definir e completar. Ademais, mediante unha proba que consistía en parafrasear un texto (explicación e xustificación), valoraron a relación entre o coñecemento léxico e a produción. Da análise efectuada, as autoras conclúen que o alumnado cun desempeño léxico superior tamén mostra maior precisión na selección de sinónimos e perífrases, na capacidade de redacción, na expresión de xuízos críticos, na adecuación á situación comunicativa e incluso menor número de erros ortográficos e mellor presentación en canto a orde e prolixidade.

Este alumnado con maior riqueza léxica presentou, por exemplo, unha utilización máis variada, máis abundante e precisa de nexos, que permiten mostrar relacións diferentes

entre ideas (*porque, como, é dicir, pero...*), así como recursos do tipo de citas textuais para fundamenta-la explicación; utilizan lexemas operativos como *expón, fala de* e outros que permiten expresións e relacións de tipo máis abstracto (*sensacións do personaxe, nun sentido metafórico, o concepto de...*); ademais, non tenden a utilizar termos como *cousa, asunto*, etc.

As autoras conclúen que, en xeral, un adestramento intensivo, sostido ó longo dun ciclo educativo, axuda a mellora-lo dominio léxico e contribúe ó logro dunha maior competencia comunicativa e, como se indicou máis arriba, a interrelación entre coñecemento léxico e textual beneficia tanto as habilidades lingüísticas como o rendemento curricular. O vocabulario constitúese nunha especie de ferramenta que posibilita a construción de pontes en canto ó significado, que sirvan para integra-los coñecementos novos con outros xa adquiridos e relacionar significacións, para dar resposta á tarefa de escribir. Ora ben, esta ensinanza do vocabulario non consiste tanto na aprendizaxe de palabras illadas como na incorporación de estratexias cognitivas e na integración de capacidades lingüísticas, como é a competencia léxica, co fin de que os escolares aprendan qué significante recolle determinado concepto, que outros significados ou acepcións ten, en que contextos se aplica, con que outras palabras se relaciona, etc. En termos xerais, trátase, por exemplo, de traballa-los termos máis importantes dun tema que se estea a estudar e utiliza-lo seu potencial semántico para explicar, exemplificar, comparar, etc. Deste modo, o léxico convértese nun instrumento fundamental para afondar no sistema conceptual dunha área.

Dentro desta concepción do vocabulario como un instrumento que permite relacionar significados e como unha variable cun alto grao de influencia na produción escrita, podería encadrarse a activación de coñecementos previos como unha tarefa básica na elaboración de textos informativos (Wray e Lewis, 2000), en tanto que nesta fase do proceso de elaboración dun texto, as ferramentas que se propoñen (tormentas de ideas, mapas conceptuais) están asentados esencialmente en palabras, no seu significado.

En resumo, parece que a investigación relativa á importancia do vocabulario na escritura non é moi abundante, pero a evidencia existente suxire que existe unha relación entre o coñecemento do vocabulario e a calidade da produción escrita, que a instrución en vocabulario en determinadas condicións permite incrementa-la calidade dos textos producidos polos suxeitos entrenados en vocabulario, cando se aprende a utiliza-lo adecuadamente como unha ferramenta que facilite a construción do tecido do texto, a integración de coñecementos novos con outros xa adquiridos e a relación de significados para responder á finalidade do texto.

Como conclusión á relevancia do vocabulario en xeral, en primeiro lugar, a investigación parece indicar que o vocabulario móstrase como unha variable relevante no que se refire ó rendemento académico. En segundo lugar, o vocabulario é unha variable que correlaciona positiva e significativamente coa intelixencia, pero, con todo, é necesario ser cautelosos na interpretación desta correlación, xa que hai investigacións segundo as cales non se pode concluír que sexa unha variable útil para predici-lo rendemento intelectual. En terceiro lugar, aínda que non foi posible achegar datos de investigación ó respecto, en base ós datos en relación coa comprensión lectora e coa expresión escrita, parece lóxico pensar que o vocabulario é unha variable relevante no tocante á lingua oral. En cuarto lugar, o vocabulario é unha das variables que mellor predi a aprendizaxe inicial da lectura, aínda non sendo a máis importante. En quinto lugar, a evidencia suxire que canto maior é o coñecemento do vocabulario, maiores son as posibilidades de comprensión dun texto, pero tamén ocorre que a lectura é unha das principais fontes, senón a principal, de aprendizaxe de vocabulario. Finalmente, os datos achegados pola investigación sobre a influencia do vocabulario na expresión escrita parecen ir no sentido de que, canto máis amplo e variado é, a súa repercusión é máis positiva na calidade da escritura, fundamentalmente cando se adquire a capacidade de empregalo como un instrumento ó servizo da mesma. Para rematar,

pódese dicir que o que parece obvio (a relevancia do vocabulario), vese sustentado en boa parte por evidencia empírica.

Capítulo 2

Estudios sobre o tamaño do vocabulario

Convén ter presente neste capítulo o comentado nos apartados 1.1 e 1.2 do capítulo 1 respecto dos conceptos de *léxico* e *vocabulario*, *palabra* e *vocábulo*, así como o que se entende por coñecer unha palabra, xa que as concepcións das que se parta ó respecto teñen unha importante influencia na estimación do tamaño do vocabulario. Convén igualmente recordar que cando se fala de vocabulario, faise referencia ó conxunto de vocábulos que, formando parte do léxico individual dunha persoa, se poñen de manifesto nunha situación determinada.

Por outro lado, cómpre sinalar que se inclúe no capítulo un apartado adicado á adquisición das primeiras palabras e o desenvolvemento do vocabulario, coa intención de proporcionar unha panorámica global do tema, que sirva para contextualizar, á vez que para complementa-la información aportada polos diferentes estudos que se exporán posteriormente.

2.1. Adquisición das primeiras palabras e desenvolvemento do vocabulario

Como se acaba de indicar, unha vez tratadas cuestións relativas ó concepto de palabra, ó que se entende por coñecer unha palabra e á relevancia do vocabulario, neste apartado abordanse sucintamente aspectos relacionados, en primeiro lugar, cos mecanismos e principios polos que se rexe a adquisición das primeiras palabras, así como as estratexias que utilizan os adultos para apoia-la aprendizaxe dos nenos e nenas; en segundo lugar, coa organización e almacenamento de palabras na memoria; en terceiro lugar, coa

evolución cuantitativa e cualitativa do vocabulario e, en cuarto lugar, coa instrución en vocabulario.

2.1.1 Mecanismos e principios de adquisición de palabras

O paso das formas foneticamente consistentes do balbucido ó que se poderían considerar palabras, prodúcese sobre os 12 meses de idade. Estas “palabras” confúndense con vocalizacións prelingüísticas e xestos, de modo que se presenta unha situación pouco clara en canto ás emisións sonoras. Parece existir unha estreita relación entre estas primeiras palabras e os xestos indicativos que se dan entre os 8 e os 10 meses, relación que determina que estas palabras sexan puramente referenciais, é dicir, etiquetas asignadas a sucesións completas de acontecementos. Por exemplo, a palabra *coche* refírese a algo así como “ven mamá, entramos ó coche, vamos de paseo, etc.” ou a palabra *boneca* a “ven papá, vestímo-la moneca, lávolle a cara e papá rise”. A palabra etiqueta situacións complexas, repetidas de diferentes formas. A súa función ven a se-la mesma que a dos xestos indicativos: comparti-la experiencia máis que o significado (Bruner, 1978, p. 206; 1986, p. 74), polo que non reflicte unha representación mental dun determinado obxecto, senón que a relación entre a palabra e o obxecto é directa.

Arredor dos 18 meses, comezan a presentar unha capacidade para a representación mental da secuencia de acontecementos ós que se refire unha “palabra” do tipo das indicadas no parágrafo anterior. En torno a ista idade tamén empezan a ser capaces de asignar esa “palabra” a unha parte da citada secuencia en forma de palabra convencional, de modo que a palabra xa denota, no sentido de que se relaciona coa representación mental dun obxecto, acción, etc. (Gopnik e Meltzoff, 1992; Nelson, 1988a, pp. 76, 97, 106). Isto é posible polo desenvolvemento cognitivo relacionado fundamentalmente coa capacidade para categorizar que se aprende mediante a repetición de experiencias concretas en contextos nos que habitualmente se producen moi poucas variantes e que se denominan *formatos*.

O progresivo desenvolvemento do vocabulario prodúcese inicialmente no contexto da interacción e intercambio comunicativo cos adultos, os cales se adaptan ó grao de coñecemento dos nenos e utilizando diferentes mecanismos (Adams, 1986), desde os máis simples como o etiquetado e reetiquetado (corrección de erros no etiquetado) a outros máis complexos como os enunciados de identificación (*Isto é...*), utilización dunha palabra coñecida polo neno con outra nova relacionada coa anterior (*É un can boxer*) ou aseveracións de inclusión (*O xílgaro é un paxaro*). Tamén exercen unha influencia neste sentido as relacións con nenos de diferentes idades. Por outra parte (Aguado, 1995, p. 92), os adultos utilizan tamén estratexias específicas de contraste como é a de substitución de términos, sendo frecuente que o término substituto sexa hipónimo do substituído, de forma que se introduce a categorización. Outras estratexias de contraste no referido a términos de tamaño son a de proliferación (habitualmente con fórmulas do tipo de *máis grande que*) e a de dimensión, consistente en empregar un término definindo unha dimensión e, seguidamente, emprega-lo valor polar desa magnitude (*menor e maior, alto e baixo, ancho e estreito*).

Os propios nenos e nenas tamén utilizan estratexias que lle aseguren a aprendizaxe do vocabulario. Así, adóitase considerar que a aprendizaxe de palabras está constrinxida por algúns mecanismos para que sexa efectiva. Unha das constricións que se adoitan aceptar na aprendizaxe de novas palabras é a do “contraste léxico” (Clark, 1988, 1990; McShane; 1980; Nelson, 1988a, p. 107, 153 e 1988b, entre outros). Pode definirse como o mecanismo por medio do que ó oír unha nova palabra se contrasta coas que xa se coñecen e asígnaselle esa palabra nova a un referente para os que non se dispón de nome. Aínda que este mecanismo non é aceptado por autores como Markman (1989, p. 187; 1994), que propón o principio de “exclusividade mutua” e en esencia non semella contradicilo anterior, a aplicación deste principio parece evidente desde a explosión denominativa que se produce en torno ós 18

meses nos monolingües e nos bilingües, que funcionan coma se tiveran un só sistema léxico, cando menos inicialmente (Genesee, 1989; Meisel e Mahlau, 1988).

Cabe concluír respecto deste mecanismo das constricións lingüísticas (Aguado, 1995, p. 86) que a partir da explosión denominativa póñense en marcha tres constricións que delimitan as suposicións que é posible facer acerca do significado dunha palabra e que interaccionan con outras fontes de información como é o contexto sintáctico (Markman, 1994). Estas constricións son, xunto coa de exclusividade mutua xa sinalada, a tendencia dos nenos e nenas a poñerlle nome á totalidade dun obxecto e non a algunha das súas partes, calidades, accións, etc., o que se ve reforzado pola mesma forma de proceder inicial dos adultos ó falar cos nenos (*can* para “move-la cola, cor, etc.”; *papá* para referirse á súa roupa; *man* por dedos, etc.); outra constrición é a tendencia a estende-los nomes a obxectos ou entidades pertencentes á mesma clase, non restrinxíndose a un elemento concreto e único (*gato* para tódolos gatos).

Clark (1995) propón unha serie de principios para explicar como resoven os aprendices o problema de conecta-los sons que escoitan cos conceptos que lle serven como guías para a construción do léxico inicial, así como para predici-la evolución do mesmo. Por unha parte, dous principios pragmáticos relacionados coa tendencia a reproducir-las formas convencionais que escoitan. Entre estes o primeiro é o da convencionalidade que se refire a que as palabras funcionan porque se acepta a convención, á conciencia de que un conxunto de sons correspóndese consistentemente cun significado particular, é dicir, que determinada palabra significa algo. “O que fai que os termos sexan convencionais é que tódolos falantes dunha comunidade están de acordo respecto de a qué clases denotan as palabras ou expresións” (p. 67). O segundo principio é o do contraste, que non é fácil desligar do de convencionalidade, xa que cando se adopta unha forma como convencional é porque se observa tamén o contraste. Segundo o principio de contraste os nenos asumen que cada

palabra significa algo diferente doutra; unha nova palabra non pode te-lo significado doutras, distintas palabras implican diferentes significados.

Por outra parte, Clark cita tres principios de procesamento, que complementan ós pragmáticos na explicación de como relacionan os aprendices os sons cos conceptos e de como evoluciona o léxico inicial. O primeiro é o de transparencia no significado, en referencia a que é máis probable que as primeiras palabras que se aprenden son aquelas das que se coñece o significado dos seus compoñentes. Por exemplo, en inglés *scissor man* en vez de *cutter*, *river bank* en vez de *riparian* ou en galego *home das flores* en vez de *xardineiro*, *coche de enfermos* por *ambulancia*, *médico de animais* por *veterinario*, etc. "Os falantes tratan de interpretar e aprender aquelas palabras que son máis transparentes, é dicir, palabras que constan de palabras base, raíces e afixos coñecidos" (p. 116). Alén disto, hai palabras cuxa existencia se explica polo feito de que é máis probable que haxa palabras para aquilo que é máis facilmente perceptible e representable; así, probablemente non se utilice a palabra *casa* ata que non sexa perceptible en tamaño reducido, en imaxes ou en xoguete. O segundo principio é o de simplicidade da forma, que alude a que "os falantes encontran máis fácil interpretar e aprender palabras canto máis simple é a súa forma, é dicir, cantos menos cambios teña a palabra base na súa construción" (p. 120). Tamén alude á tendencia a que as palabras que se adquiren en primeiro lugar adoitan ser de estrutura silábica simple ou simplificable para facilita-la súa pronunciación (*papá*, *mamá* ...). O terceiro principio é o da produtividade, que reflicte a preferencia polos termos convencionais colectivos. Este principio de produtividade enténdese no sentido de que é máis probable que inicialmente os nenos e nenas utilicen preferentemente os regulamentos de formación de palabras máis habituais no propio idioma e que, polo tanto, escoitan máis a miúdo. Tamén é máis probable que seleccionen as opcións estruturalmente posibles no propio idioma que as que violan esta estrutura. En consecuencia, no caso dun neno galego será máis probable que forme palabras

que denominan axentes ou oficios utilizando sufixos como *-eiro* porque é a forma que oírán con maior frecuencia.

A partir dos catro anos aproximadamente, os nenos van establecendo redes semánticas cada vez máis complexas, relacionando palabras e establecendo redes xerárquicas de significación, nas que os termos máis abstractos e inclusivos (palabras supraordinadas) se sitúan nos vértices e as máis concretas (subordinadas) na base. Por exemplo, sobre os cinco anos comezan a utilizar con certa frecuencia a palabra *animal* para referirse a un gato, cando ata ese momento unicamente utilizaban *gato* e comezan a diferenciar distintas variedades de gatos, coches, etc. (Anglin, 1977, p. 221). Do mesmo xeito, irán diferenciando significados ata ese momento confusos ou pouco claros para eles.

Igualmente, se ata antes dos catro anos adoitan utilizar esencialmente os adxectivos *grande/pequeno* para referirse ás dimensións dos obxectos, progresivamente empregarán termos máis específicos como *alto/baixo*, *longo/curto*, *ancho/estreito*. Isto mostra a tendencia xeral que se produce evolutivamente, consistente en comezar usando termos máis xerais, pasando paulatinamente ó uso doutros máis específicos (Aitchison, 1987, p. 95; Gentner, 1975; Pease e Berko, 1985). Entre os catro e nove anos desenvólvese a capacidade de definir palabras, de modo que inicialmente se refíren a aspectos concretos, baseadas na forma ou utilidade, por exemplo, para pasar a ser máis abstractas, utilizando sinónimos e categorizacións. Outra capacidade que se vai desenvolvendo entre os catro e dez anos e ten que unha gran influencia na adquisición do significado das palabras é a de usa-lo contexto lingüístico no que aparecen as mesmas. Con todo, hai tipos de palabras máis difíciles de aprender, como é o caso dos deícticos (pronomes *eu/el*, *isto/iso*; adverbios *aquí/alá*; verbos *traer/levar*, *vir/ir*) porque esixen unha capacidade que implica a superación do egocentrismo cognitivo.

Respecto da adquisición e do desenvolvemento do vocabulario ó longo de diferentes períodos evolutivos, da investigación de Anglin (1970, pp. 98-100) pódese concluír que:

a) Un aspecto importante é a crecente percepción das relacións sintácticas e conceptuais entre palabras. As persoas máis novas tenden a producir asociacións e a agrupar palabras seguindo o eixo sintagmático (*comer-almorzo; comer-mesa; escuridade-noite*), mentres que as máis maiores fano utilizando o eixo paradigmático, isto é, palabras que poden ser substituídas sintacticamente, incluíndo antónimos e hipónimos (*mesa-silla; escuridade-claridade; comer-beber*).

b) Prodúcese unha progresión do concreto ó abstracto. En relación co anterior está o crecemento da conciencia das relacións máis abstractas entre palabras e que se manifesta máis no modo en que poden concorrer con outras que nas posibles organizacións xerarquizadas das mesmas ou as redes que se poidan establecer. Os adultos mostran unha maior apertura en canto ás combinacións de palabras que poden facerse ou en canto a que palabras poden acompañar a que outras, o que permite a utilización de expresións metafóricas.

c) A progresiva capacidade de xeneralización permite diferencia-las clases máis extensas ás que podan pertence-las palabras. Deste modo, primeiro vese que os caraveis e xeranos son flores; que os carballos e os piñeiros son árbores; que as flores e as árbores son plantas; que as plantas son seres vivos, etc., o que incrementa a habilidade para relacionar categorías de características como as sinaladas nos dous parágrafos anteriores.

Para rematar, cabe dicir que, co tempo, a lectura chega a constituír unha das principais fontes de adquisición do vocabulario (Anderson e Nagy, 1993; Nagy, 1988; Nagy, 1997; Stenberg, 1987) e esta relación entre lingua escrita e vocabulario aparece como causal incluso cando se controla a intelixencia (Stanovich e Cuningham, 1992). Ó respecto, débese ter en conta que, por exemplo, as palabras de moderada e baixa frecuencia, as que diferencian máis entre os individuos con tamaños de vocabulario distintos, aparecen máis a miúdo en lingua escrita que na falada (Ellis, 1995).

Para explica-lo proceso de adquisición de vocabulario mediante a lectura, existen diferentes hipóteses alternativas (Ellis, 1995):

a) Hipótese forte de aprendizaxe implícita. Mantén que o significado dunha nova palabra adquirese de forma totalmente inconsciente, como resultado de abstraer de repetidas exposicións da palabra en distintos contextos.

b) Hipótese débil de aprendizaxe implícita que propón que a adquisición non é posible se, cando menos, non se é consciente de que é unha palabra nova (Schmidt, 1990).

c) Hipótese débil de aprendizaxe explícita segundo a cal os individuos son procesadores activos de información e o feito de que non se ensine o vocabulario non quere dicir que as persoas non utilicen estratexias para aprendelo.

d) Hipótese forte de aprendizaxe explícita, que mantén que a aplicación dunha serie de estratexias metacognitivas, xunto con coñecementos lingüísticos e do mundo en xeral, facilita a adquisición de vocabulario, como son reparar que a palabra non é familiar, intentar inferir-la palabra do contexto ou adquirirla definición mediante consultas a outras persoas, a un dicionario ou a vocabularios, intentar consolidala mediante repetición e utilizando estratexias de asociación ou de mediación, como nas que unha palabra coñecida serve de chave para aprender outra nova (Carlo et al., 2004; Ellis, 1995; Nagy, 1997; Wesche e Paribakht, 1994). A probabilidade de que unha palabra sexa aprendida incidentalmente, segundo autores como Swanborn e Glopper (1999), sitúase nun 15%, o que supón que serían necesarios uns oito encontros coa mesma palabra para que puidese ser aprendida nun alto grao de probabilidade, número de encontros que outros autores elevan a 10 (Wesche e Paribakht, 1994). Hai investigadores (Anderson e Nagy, 1993; Nagy, 1997) que argumentan que a aprendizaxe incidental é unha vía importante de adquisición do vocabulario. Así, a probabilidade de que un escolar de primaria ou secundaria aprenda unha palabra na primeira vez que se encontra con ela no contexto está entre 1/20 e 1/7 dependendo do tipo de texto e do espazo de tempo transcorrido entre a lectura e a avaliación do vocabulario. Non

obstante, segundo o cálculo destes autores con 25 minutos de lectura diarios leríanse 1 millón de palabras ó ano e, asumindo que cando menos un 2% sexan descoñecidas, alcanzarían a cifra de 20.000, co que 1/20 serían 1000 palabras ó ano de incremento. Con todo, a aprendizaxe por esta vía depende de factores como o volume de lectura e a porcentaxe de palabras descoñecidas dos textos, á vez que as palabras non sexan absolutamente descoñecidas para o lector, a fin de que a ganancia a partir do contexto sexa significativa (Nagy, Herman e Anderson, 1985; Stallman, 1991).

2.1.2 Organización e almacenamento das palabras

Outra cuestión relacionada coa adquisición de palabras e co curso que segue a evolución do vocabulario é a relativa á organización e almacenamento das representacións léxicas na memoria. En termos xerais, o coñecemento que calquera falante posúe sobre as palabras da súa lingua pode caracterizarse como un listado de entradas léxicas, que a modo de símil, sería como un dicionario monolingüe ó que se lle engade algunha información complementaria (Aitchison, 1987, p. 10; Valle Arroyo, 1991, p. 34). Agora ben, o símil entre o léxico mental e un dicionario da lingua non resulta totalmente adecuado, porque non parece posible imaxinar que no léxico mental as entradas aparezan por orde alfabética, nin está claro que estas entradas estean representadas de forma independente unhas das outras e, ademais, as propiedades sintácticas das palabras non figuran nos dicionarios da lingua. Cada entrada, cada palabra deste suposto dicionario incluiría (Belinchón, Rivière e Igoa, 1992, p. 372):

a) Unha representación fonolóxica semellante á pronunciación fonémica que aparece nos dicionarios de ortografías non transparentes, na que se especifican os fonemas que a compoñen, a orde en que están, o seu patrón de acentuación e, relacionada con este, a estrutura silábica.

b) Unha representación ortográfica composta por unha secuencia de letras, que é a forma que permite recoñecela cando está escrita.

c) Unha representación morfolóxica que consta de unidades subléxicas (prefixos, sufixos, etc.), que especifican as formas de derivación que se corresponden coa entrada léxica. Esta representación morfolóxica está relacionada coa estrutura derivativa da palabra e, a veces, coa categoría gramatical, xa que, por exemplo, determinados sufixos como *–dade* identifican ó substantivo (*felicidade, caridade*), outros como *–ando* a verbos (*xogando*) e outros como *–oso* (*cariñoso*) ou *–able* a adxectivos (*domesticable*), pero tamén poden transformalas en categorías gramaticais distintas (o verbo *comunicar* da o adxectivo *comunicativo*, o verbo *desexar* da o adxectivo *desexable*).

d) Unha representación sintáctica que establece a categoría gramatical da entrada e o contexto da estrutura en que aparecen.

e) Unha representación do significado, cuestión acerca da que existen principalmente tres grandes perspectivas teóricas que se comentarán deseguido.

f) Unha lista de termos relacionados coa palabra, lista máis numerosa que as chamadas a outras entradas que adoita haber nos dicionarios ordinarios da lingua e que é un dos aspectos esenciais a ter en conta na mellora da comprensión.

Como se acaba de sinalar, existen fundamentalmente tres propostas teóricas ou modelos para explicar como está representado o significado das palabras no léxico interno e na memoria, é dicir, a representación mental dos conceptos (Belinchón, Rivière e Igoa, 1992, p. 374; Valle Arroyo, 1991, pp. 63-67). Para uns autores, que se encadrarían nas teorías denominadas “modelos compositivos” de significado léxico (Katz e Fodor, 1973, p. 46; Smith, Shoben e Rips, 1974), en termos xerais, xa que existen diverxencias entre autores, o significado das palabras está representado na memoria como un agregado de trazos ou *primitivos, marcadores ou diferenciadores semánticos*, que combinándose dan lugar á representación semántica da palabra. Segundo os modelos de “redes semánticas” (Collins e

Loftus, 1975), os conceptos están organizados como redes de representación conectadas entre trazos; deste modo, cada unidade da rede representa un concepto individual, pero asociado a un conxunto de trazos semánticos primitivos. A “teoría dos prototipos” (Roch e Mervis, 1975) mantén que os conceptos son categorías difusas, representacións mentais sen límites claramente establecidos, polo que non existe un conxunto de trazos primitivos que sexa necesario e suficiente para definir adecuadamente unha categoría conceptual; por elo, a asignación dun obxecto a un concepto efectúase en base á *semellanza familiar* entre o obxecto e os exemplares máis característicos do concepto.

O recoñecemento das palabras, acceso ó léxico e identificación de palabras son expresións que se utilizan a veces como sinónimos, aínda que non o sexan exactamente (Álvarez, Alameda e Domínguez, 1999, pp. 91-92). O acceso ó léxico abarcaría os procesos dende a análise do estímulo ata a selección da estrutura adecuada no léxico mental, tendo lugar no momento en que as propiedades dunha palabra (ortográfica, fonolóxica, etc.) se encontran dispoñibles. O recoñecemento dunha palabra sería posterior, cando a persoa é consciente de que viu ou oíu a palabra, ó final do proceso.

Para explica-la activación do significado das palabras existen principalmente tres tipos de modelos (Belinchón, Rivière e Igoa, 1992, pp. 402-411), que aquí practicamente só se nomearán. O primeiro tipo son os “modelos autónomos”, que propoñen a existencia dunha serie de procesos encargados de realizar unha busca a partir da información perceptiva recibida; propugnan a independencia do lingüístico respecto a outros mecanismos cognitivos non lingüísticos, que os procesos son lineares, non simultáneos, e que as palabras están ordenadas nos arquivos mentais en función da súa frecuencia. O segundo tipo de modelos son os “modelos interactivos”, segundo os cales son varias as representacións (semánticas, pragmáticas) que interveñen xunto coa información perceptiva no recoñecemento da palabra, utilizando, ademais, a información contextual extralingüística. O terceiro tipo son os “modelos mixtos”, segundo os cales inicialmente dase un proceso autónomo de análise de

fonemas ata descubri-la forma fonolóxica, que se activa e se poñen en marcha tódalas informacións de diferentes ámbitos lingüísticos e extralingüísticos, como nos modelos interactivos, para acceder ó significado nese contexto.

Para estuda-lo proceso de recoñecemento de palabras utilízanse, entre outras, tarefas de decisión léxica, asumindo que é necesario acceder ó léxico mental para decidir se un estímulo é ou non unha palabra. Esta tarefa, como se verá, presenta unhas características semellantes á requirida na proba construída que se presenta na segunda parte desta tese e utilizada para estima-lo tamaño do vocabulario coñecido por escolares galegos.

2.1.3 Ritmo de adquisición e evolución cuantitativa e cualitativa

Existen publicacións de investigacións sobre o tamaño do vocabulario desde principio do século XX. Ora ben, como se reflicte nos datos que se ofrecerán a continuación, chama a atención o feito de que haxa bastantes diverxencias en canto ás cuantificacións feitas polos diferentes autores, así como no relativo á taxa de crecemento. A inclusión destes datos ten por obxecto unicamente mostra-lo estado en que se encontra o tema das estimacións cuantitativas, porque pode servir como un elemento máis para contextualiza-la finalidade desta tese: construír un test para estima-lo tamaño do vocabulario en lingua galega.

Estas discrepancias teñen a súa base nas diferenzas entre as distintas investigacións, en primeiro lugar, en canto á definición de que constitúe unha palabra, que contar como unha palabra; en segundo lugar, no que concirne á concepción de que significa coñecer unha palabra e, en terceiro lugar, no que se refire ás diferenzas en relación á metodoloxía utilizada na estimación. Xa se abordaron as dúas primeiras cuestións (apartados 1.1 e 1.2). A terceira, as diverxencias no relativo á metodoloxía, tratarase con maior profundidade neste capítulo 2 e no 3. Aquí será suficiente con sinalar a grandes trazos que os estudos que abranguen fundamentalmente ata os 5 ou 6 anos baséanse habitualmente en recontos de

vocabulario activo absoluto que se realizan mediante gravacións de conversas e de situacións de xogo, pero en ocasións tamén empregan tarefas de denominación de debuxos. Por outro lado, os estudos con idades máis avanzadas baséanse en estimacións de vocabulario activo ou cálculos de frecuencias e estimacións do tamaño absoluto do vocabulario. Estas últimas varían, a súa vez, en canto ó tipo de tarefas utilizadas, así como en canto ó tipo de fonte de que se extraen as palabras ou á metodoloxía de mostraxe para efectuar esta extracción de palabras da fonte. Outra cuestión a ter en conta nos estudos con idades máis avanzadas é que se producen diferenzas nas estimacións segundo o nivel socioeconómico (García Hoz, 1946; Carlo et al., 2004, por exemplo) e segundo o xénero (García Hoz, 1946; Justicia, 1983, 1985a, 1985b, 1995, entre outros), o que non se adoita especificar cando se ofrecen datos sobre o tamaño do vocabulario.

Polo que respecta á evolución cuantitativa do vocabulario, proporciónase a continuación información de diferentes investigacións con suxeitos de distintas idades que poderá servir como exemplificación das enormes discrepancias existentes, o que obriga a adoptar unha actitude cautelosa cos datos, pero tamén servirá para ofrecer unha idea da situación. Ó respecto, o desenvolvemento do léxico é lento ata aproximadamente os 20 meses, estimándose que un neno de 12 meses comprende en torno a 3 palabras diferentes, e entre os 15 e os 20 meses, unhas 20 palabras. O maior aumento proporcional comeza a producirse a partir deste momento e ata os catro anos en que a curva de incremento segue a ser acusada. Posteriormente, a curva que representa o incremento do vocabulario suavízase e, segundo o cálculo efectuado por Aguado (1999, p. 80), as desviacións estándar aumentan notablemente (Figura 2.1.), o que reflicte que as diferenzas individuais vanse incrementando.

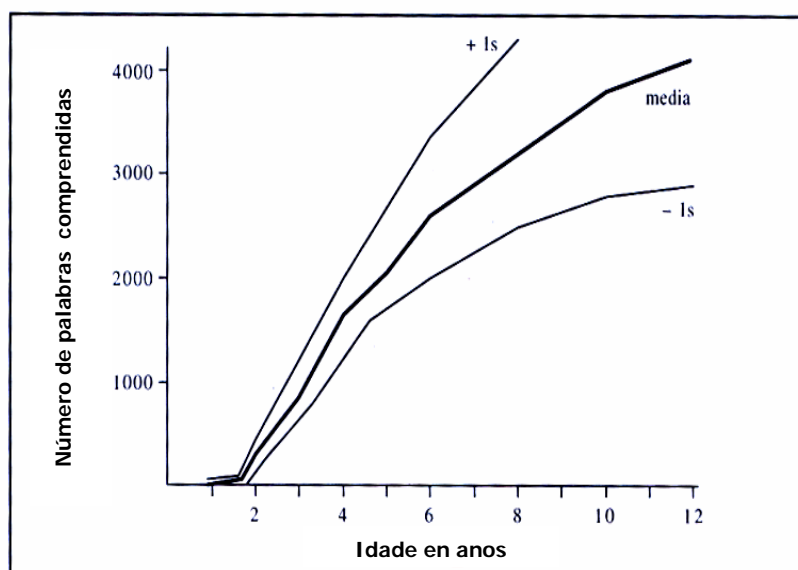


Figura 2.1.: Trazado especulativo do crecemento do léxico, baseado nos datos de Smith, 1926 (Aguado, 1999, p. 80)

Segundo Jones, Smith e Landau (1991), na primeira metade do segundo ano en que adquire as primeiras 30 palabras, aprende de 3 a 5 palabras mensuais, mentres que as seguintes miles adquírenas a razón de 10 a 20 diarias. Pero a adquisición dos diferentes tipos de palabras non se produce de forma equitativa en canto á cantidade. O vocabulario do segundo ano componse fundamentalmente de nomes nun 60%, adxectivos nun 30% e verbos nun 10% (Fletcher, 1987), pero existen variacións ó respecto entre investigacións diferentes (Dale, 1991; Reznick e Goldsmith, 1989, por exemplo). As palabras de tipo pechado (preposicións, conxuncións, etc.) aparecen máis tarde.

Durante o terceiro ano os nomes supoñen o 34% do total das palabras, no caso de nenos chilenos (Pandolfi e Herrera, 1985). Entre os nomes o 34% son de persoas, o 16% de animais e o outro 50% repártese entre varios grupos de entidades inanimadas (obxectos da casa, vehículos, etc.). Os verbos supoñen o 21% do total de palabras, sendo os máis utilizados os de acción e cambio de estado (58%). Os adxectivos representan un 8% e os pronomes e adverbios un 10% cada un deles. Por outra parte, neste período evolutivo (entre 2 e 3 anos) prodúcese un desfase entre o vocabulario que comprenden e o que poden

expresar, que é máis acusado no caso dos verbos (89% dos verbos comprendidos fronte o 42,5% dos expresados), mentres que en nomes é do 99% e do 90%, respectivamente (Peraíta, 1986).

Os datos que se obteñen en investigacións sobre o tamaño do vocabulario oral con alumnado de 1.º curso de ensinanza primaria oscilan considerablemente desde unhas 2500 palabras (Dolch, 1936; Smith, 1926) a 25.000 (Smith, 1941). Algunhas estimacións suxiren que a partir de 2 anos aprenden sobre 10 palabras novas diariamente, ata alcanzar un vocabulario en torno a 14.000 ós 6 anos (Clark, 1995, p. 13). Chall (1987) estima que sobre esta idade coñecen unhas 6000 palabras, mentres que, en España, Justicia (1983, 1985a, 1985b) ofrece unha estimación dunhas 1676 palabras para o seu vocabulario usual e un incremento anual que vai desde 335 ós sete anos a 1080 ós nove anos. En canto ás categorías dos vocábulos, segundo a investigación de Justicia, as proporcións entre seis e dez anos seguen unha pauta semellante á encontrada por Pandolfi e Herrera (1985), aínda que a metodoloxía fose absolutamente diferente. Esta investigación mostra, así mesmo, que, coa idade, descende a proporción de nomes e pronomes, mentres que aumenta a proporción do resto das categorías (Táboa 2.1.). Outros autores estiman o incremento durante a educación primaria nunhas 3000 palabras anuais (Anglin, 1993; Miller e Gildea, 1987; Nagy e Herman, 1987).

Táboa 2.1.: Porcentaxes de vocábulos por categorías (Justicia, 1985a)

Idade	Categorías				
	Nomes %	Adxectivos %	Verbos %	Adverbios %	Pronomes %
6 anos	67,84	10,86	15,24	2,15	1,90
7 anos	64,23	12,64	16,63	2,85	1,80
8 anos	63,16	14,03	17,32	2,36	1,66
9 anos	62,83	15,62	17,04	2,18	1,20
10 anos	61,43	16,41	17,71	2,29	1,03

Smith (1941) estimou o vocabulario de escolares de 13 anos nunhas 51.000 palabras. Efectuando unha extrapolación dos datos ofrecidos por García Hoz (1946) pódese estimar nunhas 22.209 a esta idade, mentres os datos de Suárez, Seisdedos e Meara (1998) ofrecen unha estimación dunhas 7800 palabras e un incremento anual dunhas 812. Para Diller (1978) a aprendizaxe anual en estudantes de educación secundaria sitúase nunhas 20.000 palabras, o que significa unha razón de 60 diarias, e para Diack (1975) esta aprendizaxe diaria é dunhas 5 palabras.

Ó remate do bacharelato, Nagy e Herman (1987) estiman que o vocabulario é de 40.000 palabras aproximadamente, de forma que se aprenderían entre 2700 e 3000 palabras novas cada ano (Nagy e Herman, 1987; Beck e McKeown, 1991) ou unhas sete palabras por día ata alcanza-las 40.000 (Cooper, 1997, p. 231). Jonhson (2000), pola súa parte, entende que probablemente alcancen as 45.000, incrementándose a razón de 10 por día como media.

As estimacións con estudantes universitarios e adultos van desde 3000 a 216.000 (Diller, 1978; Fries e Traver, 1960; Lorge e Chall, 1963). Graves (1986), nunha revisión de estudos sobre o tamaño do vocabulario feitos entre 1891 e 1960, encontrou que as estimacións para graduados universitarios oscilaban entre 19.000 e 200.000 palabras; Goulden, Nation e Read (1990) suxiren que se produce un incremento anual dunhas 1000 palabras.

Resaltouse anteriormente a existencia dunha gran diverxencia entre os datos ofrecidos por diferentes autores. Tamén se explicou sucintamente a razón pola que se incluían este tipo de datos a pesar das, en ocasións, enormes discrepancias. Estas discrepancias derivan, en boa parte, das diferenzas no que se refire ás metodoloxías empregadas na recollida de información. Se se explicase con detemento o procedemento utilizado en cada unha das investigacións, probablemente sería menor a impresión de

confusión que poda presentar a exposición e se entenderían mellor os datos, pero, por outra parte, sería excesivamente prolixa a exposición para a finalidade con que se inclúen os datos. Por este motivo, optouse por presenta-la información sen máis aclaracións, o que si se fará máis adiante cando se expoñan estudos en español e galego e na segunda parte coa achega empírica.

2.1.4 Instrución en vocabulario

En canto á instrución en vocabulario, non parece discutible que o vocabulario apréndese de distintas formas, que van desde a aprendizaxe incidental por medio da lectura e escoita dunha ampla variedade de textos ata a aprendizaxe mediante instrución directa. Pero hai autores que, na liña do comentado ó final do apartado 2.1.1., poñen máis enfase na importancia duns mecanismos sobre outros, como é o papel da lectura, aínda que sexa necesario aprender a usa-lo contexto para poder empregalo na construción do significado (Anderson e Nagy, 1993; Cooper, 1997; Gilabert, 1995; Gilabert e Vidal-Abarca, 1993; Graves, 1987; Nagy, 1997; Nagy e Herman, 1987; Sternberg, 1987; Wesche e Paribakht, 1994).

Unha instrución eficaz debería basearse en elementos como a toma de conciencia das palabras, na lectura e escritura, na aprendizaxe de estratexias que favorezan a autonomía para inferir significados a partir do contexto e, de forma limitada, na ensinanza directa de vocabulario e de habilidades relacionadas.

A maior parte do vocabulario apréndese sen unha ensinanza directa, pero si que é necesario que os aprendices sexan conscientes das palabras que non coñecen e da necesidade de aprendelas para poder comprender. A apropiación das palabras ten lugar cando unha persoa é capaz de conectalas cos coñecementos previos, desenvolvendo redes de relacións mediante experiencias repetidas con elas na fala, na lectura e na escritura, e mediante actividades específicas dirixidas a favorecer estas relacións (elixir nun xornal

palabras que interesen, discutir sobre o significado de palabras a partir de lecturas en voz alta, elaborar arquivos de palabras, empregar mapas de palabras, etc.).

Un terceiro elemento que favorece a mellora do vocabulario é a ensinanza de estratexias que faciliten o establecemento de inferencias do significado de palabras polo contexto (Gilbert, 1995; Gilbert e Vidal-Abarca, 1993; Graves, 1987; Nagy, 1997; Sternberg, 1987). Estas inferencias aséntanse en tres procesos básicos: a codificación selectiva (separación da información relevante da irrelevante en relación coa palabra a aprender), combinación selectiva (combina-los trazos relevantes e integra-la información seleccionada para elaborar unha definición correspondente á palabra descoñecida) e comparación selectiva (relacionar e compara-la nova información coa xa coñecida).

Respecto do cuarto elemento citado, a utilización da ensinanza directa de palabras e de habilidades relacionadas (claves contextuais, prefixos, sufixos ou dicionario), os resultados da investigación son contradictorios (Beck e McKeown, 1991; Graves, 1987), polo que se pode pensar que este tipo de intervención ten un valor limitado. A pesar desta evidencia segue a ser un procedemento empregado amplamente polos docentes, probablemente porque pensan que o significado dun texto reside no significado individual das palabras, porque cren que a ensinanza de palabras antes de ler un texto mellora a súa comprensión ou porque non coñecen outros procedementos diferentes ós de ensinar directamente palabras. Segundo Cooper (1997, p. 233), da revisión de diversos estudos, pode deducirse que este tipo de instrución resulta máis efectiva cando ten lugar atendendo a unha serie de condicións: son ensinadas unhas poucas palabras centrais no contido do texto, sexa narrativo ou informativo, pero non a ensinanza de palabras ó azar; as palabras son ensinadas en contextos significativos onde expresen o significado relevante para o contexto; a instrución de novas palabras intégrase coa activación e desenvolvemento do coñecemento previo dos aprendices; a ensinanza lévase a cabo ofrecéndolle ós escolares información rica e variada sobre as mesmas, á vez que as palabras están relacionadas entre si e coa

experiencia do alumnado; finalmente, cando os escolares teñen contacto coa mesma palabra en multitude de ocasións e o alumnado participa activamente no proceso de aprendizaxe, verbalizando o que aprende e relacionándoo coa súa experiencia.

En síntese, no que atinxe á instrución, aínda que a lectura e escritura sexan as principais vías por medio das cales se incrementa o vocabulario, é posible utilizar estratexias que favorezan a aprendizaxe do significado de palabras de forma autónoma, estratexias que deberían sustentarse no establecemento de relacións entre o novo vocabulario e o coñecemento previo, na produción de textos polos aprendices e na inferencia do significado. Por outro lado, é preciso adoptar decisións no relativo á selección de palabras (excluindo aquelas que están definidas no texto e as que se poden deducir por medio do coñecemento de sufixos, prefixos e a súa combinación co coñecemento de palabras base), pero tamén adoptar decisións en relación con cando ensinar estas palabras e en relación coa estratexia ou estratexias a utilizar na ensinanza (mapas de palabras, análise das características semánticas, esquemas de relacións xerárquicas, etc.).

Deseguido efectúase un resumo de todo o apartado 2.1. En primeiro lugar, a adquisición do vocabulario aséntase nunha serie de piares como son: a interacción con adultos e con coetáneos; a aplicación de construcións lingüísticas como son a do “contraste léxico”, a tendencia a poñerlle nome á totalidade do obxecto e a xeneralización dos nomes a obxectos ou entidades da mesma clase; o progresivo establecemento de redes semánticas; o incremento constante da influencia da lectura, que se explica mediante diferentes hipóteses, como as hipóteses débil e forte de aprendizaxe implícita e as hipótese débil e forte de aprendizaxe explícita. En segundo lugar, as palabras aprendidas almacénanse na memoria organizadas de forma semellante a un dicionario da lingua. Hai principalmente tres modelos explicativos da representación do significado das palabras almacenadas: os composiciónais, os de redes semánticas e a teoría dos prototipos. En terceiro lugar, en canto á explicación de

como se activa o significado, existen tamén fundamentalmente tres tipos de modelos: os autónomos, os interactivos e os mixtos. Polo que respecta á evolución cuantitativa do vocabulario, os datos que aporta a investigación son moi dispares, debido ás diferenzas en canto ós procedementos usados para a estimación. As seguintes palabras de Pease e Berko (1985) reflicten ben a idea xeral en relación coa adquisición das primeiras palabras e o desenvolvemento do vocabulario:

O incremento do vocabulario consiste na adquisición de novas palabras e novos significados e no establecemento de relacións entre elas. Para aprende-lo significado dunha nova palabra o neno debe aprender como ten agrupado e denominado a súa comunidade lingüística os diferentes aspectos da experiencia; é a convención social e non a experiencia individual ou o mundo dos obxectos o que define as palabras. O desenvolvemento semántico caracterízase tanto pola adición de novos elementos de significado como pola reorganización dos elementos coñecidos (p. 130).

2.2 Tipos de estudos sobre o tamaño do vocabulario

A valoración do coñecemento do vocabulario ten sido abordada en base a dúas perspectivas fundamentalmente: tamaño e profundidade. Son dous enfoques que se teñen considerado complementarios e dalgún modo incompatibles, presentando cada un distintas vantaxes e limitacións.

A perspectiva que se centra no tamaño do vocabulario será a que interesa directamente nesta tese. Estas medidas do tamaño teñen resultado eficaces como indicadores para situar alumnado no nivel máis adecuado de aprendizaxe dunha lingua estranxeira, por exemplo, ou na valoración do incremento do número de palabras coñecidas

(Milton e Meara, 1995). Entre os estudos deste tipo, destacan tres perspectivas, asociadas á utilización de tres tipos de procedementos para estudalo (Justicia, 1985b, 1993, 1995).

2.2.1 Recontos do vocabulario activo absoluto

Neste apartado farase referencia a estudos que algúns autores denominan recontos absolutos de vocabulario (Justicia, 1985b, 1993, 1995). Son as investigacións máis numerosas e teñen como fin determina-lo número total de palabras emitidas oralmente por un individuo concreto ou por mostras reducidas ó longo dun curto período de tempo, fundamentalmente durante os primeiros anos de vida.

Existen tanto estudos que se poden denominar biográficos coma tamén con grupos. En ámbolos dous casos se traballa con nenos e nenas de idades que se corresponden coa etapa do ensino infantil. A tarefa é factible en tanto que o tamaño do vocabulario destes nenos e nenas está dentro duns límites manexables. A grandes trazos, a metodoloxía utilizada baséase esencialmente na observación en situacións de carácter máis natural, ou menos artificiais que as valoracións mediante tests. Nos estudos de produción baséase en gravacións de conversas que posteriormente se transcriben, pero tamén se teñen usado probas de denominación de imaxes, ou, nos estudos de comprensión, tarefas de recoñecemento, nas que a resposta consiste en indica-la figura cando é nomeada pola persoa que leva a cabo a exploración. As máis frecuentes efectúan un rexistro sistemático da produción verbal nun espazo de tempo e describen o vocabulario cualitativa e cuantitativamente. Recompílanse gran cantidade de datos sobre a adquisición do vocabulario desde que o neno articula as primeiras palabras ata o cuarto ou quinto ano e, ademais de medicións do tamaño, levan a cabo análises doutro tipo de variables relacionadas coa evolución do vocabulario (que tipo de palabras se adquiren e en que proporcións, como evolucionan determinadas formas, as diferenzas entre suxeitos de distintas características, a influencia de diversas variables, a estrutura de diferentes tipos de frases, etc.). As mostras

de fala máis habituais recollen entre 50 e 100 expresións, pero as que ofrecen máis e mellor información utilizan de 300 a 800 ou incluso máis.

Estas investigacións presentan unha serie de dificultades, porque, en primeiro lugar, na medida en que se observa un rápido crecemento do vocabulario en relación coa idade, o reconto de palabras do léxico infantil vese dificultado, converténdose en practicamente inviable máis alá dos tres primeiros anos. En segundo lugar, resulta imposible empregar mostras representativas de poboación, polo que non son de gran utilidade para establecer normas xerais sobre os patróns de desenvolvemento que pode segui-lo vocabulario. En terceiro lugar, resulta moi difícil establecer comparacións entre investigacións debido ós diferentes procedementos usados na recollida de datos. En cuarto lugar, utilizan gravacións de fala espontánea (vocabulario produtivo) sen explorar outros niveis de coñecemento. Prodúcese, así mesmo, unha gran dispersión entre suxeitos en función de variables como a situación, o tema tratado ou a familiaridade do observador.

Un dos estudos máis coñecidos deste tipo levouno a cabo M. E. Smith (1926) utilizando un test que requiría a evocación de nomes a partir de debuxos ou, no caso de que non fose posible a evocación, sinala-la imaxe que representaba a palabra que emitía a persoa adulta. A estimación de Smith (Táboa 2.2., tomada de Dale, 1976, p. 174) probablemente signifique unha subestimación debido a que o universo de 10.000 palabras de que partiu era moi limitado. Así, estimou un vocabulario de 2562 palabras para os 6 anos, mentres que empregando metodoloxías distintas, con frecuencia, os datos que se achegan na investigación reflicten un maior tamaño do vocabulario nesta idade como se viu no punto 2.1.3.

No que se refire á variabilidade individual, Stoel-Gammon e Cooper (1984) mostraron como das 50 primeiras palabras emitidas por dous nenos e unha nena, soamente 8 eran comúns ós tres suxeitos, o que é un indicador da importante influencia do medio ambiente na concepción do mundo e o seu reflexo no vocabulario. Lieven, Pine e Dresner (1992), nun

estudo con 12 nenos e nenas ata 3 anos, comprobaron que os suxeitos acadaban as 100 primeiras palabras entre 1 ano e 3 meses e 2 anos e 3 meses e que existían importantes diferenzas nas porcentaxes dos distintos tipos de palabras que se incluían entre esas 100 primeiras. De forma semellante, Serra e Serrat (1996) nun estudo con 8 nenos entre 1 e 4 anos, encontraron que entre as 100 primeiras palabras que emitiron había moi poucas comúns a tódolos nenos. Polo que se refire ó tipo de palabras utilizadas, entre as 50 primeiras emitidas existían diferenzas entre os suxeitos, que se agruparon en dous conxuntos, os denominados referenciais e os denominados expresivos: entre as producidas polos primeiros, un 50% ou máis eran nomes, mentres que no caso dos segundos a maior porcentaxe correspondíalle ós verbos e ós adxectivos.

Táboa 2.2.: Desenvolvemento do vocabulario en número de palabras, segundo Smith (1926)

Idade	Número de palabras	Incremento
8 meses	0	—
10 meses	1	1
12 meses	3	2
15 meses	19	16
19 meses	22	3
21 meses	118	96
2 anos	272	154
2 anos e 6 meses	446	174
3 anos	896	450
3 anos e 6 meses	1222	326
4 anos	1540	318
4 anos e 6 meses	1870	330
5 anos	2072	202
5 anos e 6 meses	2289	217
6 anos	2562	273

2.2.2 Estimacións do vocabulario absoluto coñecido

Entre estes estudos están os que se corresponden cos que algúns autores denominan vocabularios de recoñecemento (Justicia, 1985b, 1993, 1995). Utilízanse co obxecto de salva-las dificultades que presentan os recontos absolutos. As estimacións son unha extrapolación a partir das palabras coñecidas dunha lista de palabras presentada ós suxeitos avaliados. Estímase o tamaño do vocabulario mediante probas que se constrúen principalmente utilizando vocábulos extraídos do vocabulario estándar dunha lingua tal como ven recollido nos dicionarios. A tarefa consiste en someter a valoración unha lista de palabras mediante procedementos distintos como, por exemplo, relacionar sinónimos, completar frases (García Hoz, 1946, 1977, entre outros) ou indicar se se coñece o significado das palabras do listado (por exemplo, Suárez, Seisdedos e Meara, 1998). Da proporción entre o número de palabras da lista sinaladas como coñecidas por un suxeito, o número de palabras correctamente relacionadas ou correctamente colocadas nunha frase e o número de palabras que constitúen a fonte orixinal da que se seleccionou a lista, obtéñense uns resultados que permiten efectuar un cálculo do número de palabras que ese suxeito coñece da fonte orixinal. Unha vez efectuada a estimación do número de palabras que coñece un individuo, é posible situar a este individuo en relación co termo medio do vocabulario recoñecido na proba polo grupo normativo da idade a que pertenza.

Entre as desvantaxes que presentan sinálanse o feito de que se valora esencialmente o vocabulario pasivo e a considerable dispersión dos resultados nos diferentes traballos. Esta dispersión é atribuíble á variabilidade no número de preguntas que conteñen as probas, ó diferente grao de dificultade das palabras, á diversidade en canto á extensión da fonte de procedencia da que se extraen as palabras, ós distintos criterios para selecciona-las palabras e ás diferenzas entre o vocabulario infantil e o que se adopta como referencia.

2.2.3 Estimacións do vocabulario activo coñecido

Estes estudos correspóndese cos que certos autores denominan cálculos de frecuencias (Justicia, 1985b, 1993, 1995). As estimacións consisten en cálculos inferenciais baseados na relación *type/token* ou tipo/figuración, a porcentaxe de veces que aparece a mesma palabra en relación co número total de palabras emitidas ou analizadas. *Type* ou tipo fai referencia ás palabras diferentes que aparecen nunha mostra de fala ou escritura, mentres que figuración (*token*) refírese ó número total de palabras que aparecen nesa mostra, o número total de ocorrencias, de palabras, contabilizando as que se repiten e as diferentes formas flexionadas. Son recontos de vocabulario activo, mostras de léxico (infantil), que ofrecen información sobre cales son as palabras utilizadas, a súa frecuencia e o contexto en que se producen. O termo medio de palabras que resultan por idade é sensiblemente inferior ó que proporcionan as estimacións de vocabulario absoluto coñecido ou vocabularios de recoñecemento, como cabería esperar.

Segundo Justicia (1993), algúns dos defectos deste tipo de procedementos son os mesmos que os sinalados para os recontos absolutos, en tanto que as frecuencias refírense a palabras emitidas por un reducido número de suxeitos.

Das tres grandes perspectivas no estudo do tamaño do vocabulario que se acaban de comentar, a segunda, estimación do vocabulario absoluto coñecido (2.2.2), é a que se corresponde co propósito desta tese. Como se irá precisando, dentro desta perspectiva, optouse polo procedemento baseado na tarefa de sinalar se se coñece o significado de cada palabra dunha lista que se presenta.

2.3 Breve descrición de estudos de vocabulario en castelán e galego

Posto que o obxectivo desta tese é a construción dunha proba que permita efectuar estimacións do tamaño do vocabulario e, con motivo de ofrecer unha visión xeral sobre esta cuestión da estimación do tamaño do vocabulario, neste apartado, a modo de

exemplificación, describíranse brevemente algúns estudos publicados en español e galego, nos que se utilizaron distintos tipos de probas coa finalidade de efectuar as devanditas estimacións.

2.3.1 García Hoz (1946)

Un dos iniciadores deste tipo de investigacións foi García Hoz, que en 1946 publicou *Evolución cuantitativa del vocabulario en escolares de 9 a 18 años*, que é un estudo do tipo dos sinalados no punto 2.2.2 (Estimación do vocabulario absoluto coñecido).

A mostra estaba composta por 2774 escolares entre 9 e 18 anos, dos que, como mínimo había 50 de cada sexo e das clases sociais baixa e media por idade, de modo que, cando menos, había 200 suxeitos de cada idade ata os 14 anos, incluíndo esta idade. A partir dos 14 anos o número de escolares por idade é inferior porque, na época en que se realizou o estudo, os suxeitos de clase baixa habitualmente abandonaban a escola nesa idade e o autor decidiu non incluír ós escolares desta clase social na investigación.

Utilizou unha proba configurada en cinco partes, catro compostas por 20 ítems e unha por 4 ítems, que se corresponden con catro formas diferentes de presentar as palabras para comprobar se as coñecían. Describírase máis pormenorizadamente no capítulo 3, punto 3.1.2, pero aquí pódese adiantar que na primeira parte os ítems esixen que se sinala o sinónimo dunha palabra estímulo entre un grupo delas; na segunda, trátase de seleccionar unha palabra que serve para completar unha frase coa que se define un término en forma de adiviña; na terceira parte, a esixencia consiste en asociar unha serie de palabras con outras dun grupo dado delas; na cuarta, é preciso elixir unha palabra para completar frases tipo cloze e na quinta parte a tarefa é a mesma que na segunda.

Para a elaboración da proba, o autor baseouse no dicionario da RAE de 1939, que constaba de 48.300 palabras sen contar as consideradas rexionalismos, os términos técnicos, os nomes de letras e os artigos, pronomes, preposicións, conxuncións e interxeccións.

Seleccionou 2714 palabras polo procedemento de catas, concretamente aquelas que encabezan as columnas de cada páxina do dicionario, e agrupou a mostra en 27 series de 100 palabras cada unha, suprimindo 14 palabras. Para a investigación utilizou as 100 palabras da primeira serie unicamente.

O autor achega datos en medias porcentuais sobre o vocabulario coñecido nas distintas idades polos escolares, así como sobre as diferenzas entre as medias e a significación das mesmas en función da idade, clase social e do sexo. Multiplicando a media dun grupo ou a puntuación dunha persoa en concreto por 48.300 e dividindo por 100, é posible obter a estimación do número de palabras coñecidas. A táboa 2.3., que o autor usa para presenta-los datos en función da idade, pode valer tamén como resumo dos resultados globais.

Táboa 2.3.: Porcentaxe de palabras que coñecen os escolares

Idade	Nenos		Nenas	
	Clase popular	Clase media	Clase popular	Clase media
9 anos	23,35	30,83	22,77	28,25
10 anos	28,62	42,41	24,33	32,56
11 anos	36,27	41,27	28,68	40,48
12 anos	37,40	44,58	35,94	42,70
13 anos	41,38	52,06	40,11	50,38
14 anos	42,34	53,45	44,59	55,66
15 anos		59,25		59,15
16 anos		64,34		61,39
17 anos		65,68		62,74
18 anos		66,47		63,55

Un par de exemplos poden ser útiles para mostra-lo procedemento utilizado para estima-lo vocabulario coñecido. Un neno de 11 anos pertencente á clase media coñecería

19.933 palabras ($41,27 \times 48.300 : 100 = 19.933$), mentres que un neno da mesma idade procedente da clase popular coñecería 17.518 palabras ($36,27 \times 48.300 : 100 = 17.518$).

2.3.2 Justicia (1985a, 1985b, 1995)

No marco dos estudos de cálculos de frecuencias, Justicia investigou o vocabulario de nenos entre 6 e 10 anos de medio rural e urbano en catro provincias andaluzas (Almería, Granada, Xaén e Málaga). Posteriormente, en 1995 foi ampliada a investigación en canto ó rango da idade ata os 13 anos. Tratou de abordar dous obxectivos complementarios. O primeiro era determina-lo vocabulario usual dos escolares a partir da análise da súa produción verbal en base a textos escritos. O segundo obxectivo era indaga-la relación do vocabulario usual coa idade ou nivel de escolarización, co sexo, coa zona de procedencia (rural/urbana) e coa clase social. A mostra estivo composta inicialmente por 2166 escolares (ampliada a 3402 en 1995), igualados en número no relativo a sexo e divididos en estratos en base á idade, zona e provincia (a clase social estableceuse con posterioridade á estratificación da mostra).

Para recompila-las palabras utilizou dúas probas de papel e lapis. A primeira, da que proceden as tres cuartas partes do vocabulario obtido, consistiu nunha composición sobre un tema libre a desenvolver durante 20 minutos. A segunda foi un test de asociación libre de palabras a realizar en 4 minutos.

Esta investigación correspóndense coas estimacións de vocabulario activo coñecido citadas anteriormente. A metodoloxía empregada é complexa (1985a, pp. 31-34; 1995, pp. 10-24) e non se vai a tratar aquí, pero basicamente, como se explicou no punto 1.2.3, consiste en contabiliza-los *tokens* ou formas de mostras de escritura e calcula-los *types* ou vocábulos, que resultan unha vez eliminadas as formas que se repiten total ou parcialmente. O procedemento que se utiliza neste tipo de estudos permite detectar un incremento do vocabulario coa idade, así como diferenzas en función do sexo ou a clase social, pero non

efectuar unha estimación directa do tamaño do vocabulario. En primeiro lugar, porque se centra nos vocábulos usados por grupos de individuos e na súa frecuencia. En segundo lugar, porque se parte de estimacións obtidas por outros autores para orienta-lo recuento de *tokens*, co fin de conseguir un número de *types* semellante ó logrado polos investigadores que serven de referencia (neste caso, dos datos proporcionados por Averril, 1956). Con todo, inclúese esta investigación entre os estudos sobre tamaño do vocabulario, porque os datos que proporciona o autor, fundamentalmente nas publicacións de 1985, fan referencia ós vocábulos utilizados polos grupos con que traballou, aínda que non se podan entender o número destes vocábulos como o tamaño do vocabulario de individuos determinados. Coa finalidade de proporciona-los datos de maior significación, a costa da claridade expositiva, neste punto farase referencia ós datos globais que figuran na publicación de 1995, pero cando se analicen por idades, sexo, nivel socioeconómico e zona xeográfica e se aluda a táboas, a referencia serán as publicacións de 1985, agás nos casos que se indique expresamente. Adoptouse esta decisión debido a que na publicación de 1995 figura información sobre o número total de *tokens* que utiliza cada grupo e non sobre o número de tipos ou vocábulos (palabras diferentes que utilizan). Desta forma, os datos perderían sentido para a finalidade con que se expoñen aquí.

Como se indicou no punto 1.2.3 do capítulo 1, Justicia contabilizou 528.544 ocorrencias de palabras (*tokens*). Do recuento obtivo 8937 vocábulos, que constitúen o vocabulario usual, no que a frecuencia media de cada vocábulo foi de 59,14 ($528.544/8937=59,14$), resultado de dividir o número total de ocorrencias (*tokens*) polo número de vocábulos (*types*).

En síntese (Táboa 2.4.), os datos que achega mostran que se produce un incremento do vocabulario usual desde 1676 vocábulos no grupo de seis anos a 4369 no grupo de 10 anos (1985a, 1985b) e 7731 no grupo de 11 a 13 anos (1995). En canto ó sexo (Táboa 2.4.), as nenas emiten un maior número de *tokens*, pero a diversidade léxica é maior no

caso dos nenos, xa que emiten un maior número de vocábulos distintos (1985a, 1985b). Non obstante, segundo os datos de 1995, a medida que se incrementa a idade as diferenzas das medias relativas ó sexo tenden incluso a invertirse en termos globais, se ben son de escasa importancia (Táboa 2.6.). No relativo á clase social, os datos da investigación de 1985 (Táboa 2.7.) indican que o vocabulario dos escolares de clase social alta é maior que o dos que pertencen á clase media e o vocabulario destes maior que o dos procedentes de clase baixa. Ademais, os datos que figuran na publicación de 1995 (Táboa 2.8.), tamén permiten sustenta-la idea de que os escolares de clase alta, cando menos, utilizan un maior número de palabras/*tokens* (en termos de puntuacións medias) que os de clase media e estes que os de clase baixa. No que se refire á influencia da zona, parece que é maior o vocabulario do alumnado de zonas urbanas (Táboa 2.9.).

Táboa 2.4.: Número de vocábulos que constitúen o vocabulario usual por idades

IDADE	Número de escolares	Nº total de Vocábulos	Frecuencia (<i>tokens/types</i>)	Incremento
6 anos	402	1676	14,22	
7 anos	326	2002	13,82	335
8 anos	380	2587	15,68	585
9 anos	510	3667	20,11	1080
10 anos	548	4369	21,32	702
11-13 anos	1236	7731	34,88	

Táboa 2.5.: Número de vocábulos que constitúen o vocabulario usual por sexo

Sexo	Número de escolares	Nº total de Vocábulos	Frecuencia (<i>tokens/types</i>)	Incremento
Nenos	1086	4689	26,18	478
Nenas	1080	4221	32,26	

Táboa 2.6.: Media de palabras ou *tokens* utilizados en función do sexo

Clase social	Número de escolares	Nº total de <i>tokens</i>	Media	Incremento
Nenos	1704	258.430	151,66	
Nenas	1698	270.114	159,08	7,42

Táboa 2.7.: Número de vocábulos que constitúen o vocabulario usual por clase social

Clase social	Número de escolares	Nº total de Vocábulos	Frecuencia (<i>tokens/types</i>)	Incremento
Alta	174	4867	9,55	1386
Media	418	3481	14,83	1160
Baixa	1574	2321	38,05	

Táboa 2.8.: Media de palabras ou *tokens* utilizados en función da clase social

Clase social	Número de escolares	Nº total de <i>tokens</i>	Media	Incremento
Alta	313	56.368	180,09	23,33
Media	585	91.705	156,76	4,81
Baixa	2504	380.471	151,95	

Táboa 2.9.: Número de vocábulos que constitúen o vocabulario usual por zona xeográfica

Zona	Número de escolares	Nº total de Vocábulos	Frecuencia (<i>tokens/types</i>)	Incremento
Urbana	850	5005	32,50	1252
Rural	1316	3753	25,67	

2.3.3 Casanova e Rivera (1989)

Estas autoras propuxéronse como obxectivos ofrecer un vocabulario básico para o alumnado español dos tres ciclos de Educación Xeral Básica, comproba-lo seu coñecemento e comparalo co vocabulario activo do alumnado mencionado. No tratan, polo tanto, de estima-lo vocabulario dos suxeitos, senón comprobar se coñecían o vocabulario básico proposto polas autoras.

Expúxose no punto 1.2.3 do capítulo 1 que as autoras, primeiramente, seleccionaron do dicionario da RAE (1970) por medio de xuíces un vocabulario básico composto por 10.017 vocábulos. Para comproba-lo grao de coñecemento do vocabulario pasivo por parte do alumnado dos tres ciclos de EXB, construíron unha proba a partir do vocabulario básico mediante un procedemento de mostraxe aleatoria sistemática. Quedaron seleccionados 350 vocábulos para segundo e quinto e 525 para o oitavo. Con estes vocábulos elaboraron 35 fichas nas que, mediante catro tarefas, se avalían 10 vocábulos en cada ficha en segundo e quinto e 15 vocábulos en cada ficha en oitavo.

Como se acaba de indicar, para segundo e quinto, a proba consiste en 35 fichas, en cada unha das cales avalíanse 10 palabras mediante catro tarefas. A primeira tarefa consiste en transformar unha frase que contén un dos 10 vocábulos conservando o seu significado. Asígnaselle unha puntuación ponderada de 2,5 (25% do total) e unha puntuación bruta de 1 punto. A segunda tarefa esixe que expliquen o significado dun vocábulo subliñado no contexto dunha frase, asignándolle unha puntuación ponderada de 2,5 puntos (25% do total) e unha puntuación bruta de 1 punto. En canto á terceira tarefa, en 2.º curso consiste en completar tres frases utilizando un dos tres vocábulos que se propoñen para cada frase (un correcto e dúas alternativas incorrectas ou *distractores*), mentres que en 5.º curso consiste en explica-lo significado de tres vocábulos estímulo. Asígnaselle unha puntuación ponderada de 2,5 puntos (0,83 a cada unha das tres frases, que representa o 25% do total) e unha puntuación bruta de 3 puntos. Finalmente, a cuarta tarefa consiste en emparellar

cinco vocábulos con cinco definiciónís. Recibe unha puntuación ponderada de 2,5 puntos (0,5 por cada palabra) e unha puntuación bruta de 5 puntos (un por palabra).

Para o alumnado de 8.º curso (2.º de ESO actualmente) a proba está composta por 35 fichas, en cada unha das cales avalíanse 15 palabras. A primeira tarefa consta de tres frases, inseríndose en cada unha delas un vocábulo subliñado dos 15 da ficha. Os escolares deben explica-lo seu significado (a puntuación máxima é de 3 puntos). A segunda tarefa contén tres vocábulos que aparecen illadamente, dos que deben de dar un sinónimo ou expresión sinónima de cada un (a puntuación máxima é de 3 puntos). Na terceira tarefa deben dar unha definición de catro vocábulos (a puntuación máxima é de 4 puntos). A cuarta tarefa consiste en relacionar cinco vocábulos con cinco definiciónís (a puntuación máxima é de 5 puntos).

Administráronlle as probas a unha mostra de 8714 escolares de EXB, obtida das diferentes comunidades autónomas do estado español: 2895 de segundo curso, 3228 de quinto e 2591 de oitavo, que se corresponde con 2.º de ESO actual. En 2.º e 5.º tiveron en conta a distribución xeográfica dos centros, as súas características (tamaño e localización) e réxime xurídico (públicos e privados). En 8.º os centros seleccionados formaban parte dos que experimentaban a reforma do Ciclo Superior.

Presentan as puntuacións medias brutas e ponderadas (Táboa 2.10.), desdobrando estas puntuacións en 2.º e 5.º, respectivamente conforme non se tiveran en conta as fichas nulas (efectuando o cálculo en base ás puntuacións totais e ó número total de fichas completadas) ou téndoas en conta (deducindo as fichas nulas do número total para efectuarlo cálculo). Pódese observar na táboa que o coñecemento do vocabulario avaliado é moi alto, xa que as puntuacións en segundo e quinto de EXB son superiores a 8, cando a máxima puntuación sería 10, e no caso de oitavo curso é superior a 13 dun máximo de 15. Con elo, cabe pensar que o tamaño do vocabulario pasivo destes escolares é maior que o da poboación de vocábulos do que se extraeron as mostras (o vocabulario básico).

Táboa 2. 10. : Puntuacións medias brutas e ponderadas do vocabulario pasivo coñecido, segundo Casanova e Rivera (1989)

	Puntuación Bruta		Puntuación Ponderada	
	*	**	*	**
2º EXB	8,65	8,07	8,21	9,05
5º EXB	9,42	9,53	9,16	9,38
8º EXB	13,40			

* Puntuacións medias sen ter en conta as fichas nulas (cálculo da media en base ás puntuacións totais e ó numero de fichas completadas).

** Puntuacións medias tendo en conta as fichas nulas (deducíndoas do número total para efectua-lo cálculo).

Para tratar de indaga-lo coñecemento do vocabulario activo, utilizaron dúas probas escritas que lle aplicaron á mesma mostra empregada para comproba-lo coñecemento pasivo do vocabulario básico. A primeira denominárona de escritura espontánea (evocación libre de palabras por escrito) e a segunda proba denominárona de escritura suxerida (evocación de palabras sobre un tema determinado, indicado en cada caso). Mediante estas probas obtiveron as palabras máis familiares para o alumnado (escritura espontánea) e máis coñecidas sobre temas determinados (escritura suxerida). Con elo, conseguiron unha visión do vocabulario que usaban os escolares habitualmente na expresión escrita. O corpus recompilado a partir das producións de tódolos escolares consta de 266.729 *tokens* ou número total de palabras sen elimina-las que se repiten (78.967 en segundo curso, 95.599 en quinto e 92.163 en oitavo). O termo medio de palabras emitidas por cada escolar foi de 27,27 en Ciclo Inicial, 29,61 en Ciclo Medio e 35,57 en Ciclo Superior. O vocabulario activo obtido a partir do corpus total está formado por 18.881 vocábulos e expresións diferentes (*auga osixenada, cabeza nuclear, avoa materna*, etc.): 8185 en segundo curso , 11.329 en quinto, 15.016 en oitavo, que representan correlativamente o 43,35%, o 60% e o 79,52% dos 18.881 vocábulos. Da comparación entre o vocabulario activo e o vocabulario básico

proposto polas propias autoras (10.017 palabras), observaron que os escolares utilizaron o 71% dos vocábulos considerados básicos.

De todo o exposto, en primeiro lugar, respecto do vocabulario pasivo dos escolares avaliados, pódese concluír que os vocabularios básicos de cada ciclo son unha estimación baixa do pasivo, porque os escolares demostraron que coñecían unha altísima porcentaxe dos vocábulos (Táboa 2.10.) e, ademais, pódese supoñer que con seguridade coñecerán máis vocábulos non incluídos nos vocabularios básicos. É dicir, que o alumnado coñecerá un vocabulario máis amplo que as 10.017 palabras incluídas no vocabulario básico. En segundo lugar, respecto do vocabulario activo, cabe concluír que se aproxima ó vocabulario considerado básico, posto que o alumnado avaliado utilizou en torno ó 71% dese vocabulario, pero, ademais, tamén neste caso son capaces de utilizar vocábulos non incluídos no vocabulario básico. En terceiro lugar, pódese deducir que o coñecemento pasivo do vocabulario básico é maior que o coñecemento activo. En cuarto lugar, parece que os xuíces son sensibles en relación co coñecemento do vocabulario que poden ter os escolares.

2.3.4. Esteban (1997)

Este autor levou a cabo en Galicia un estudo con alumnado de 2.º e 5.º de Educación Primaria e 8.º de EXB (actualmente 2.º de ESO) da provincia de Ourense, no que basicamente tratou de medi-los niveis de adquisición do vocabulario pasivo e efectuar unha comparación entre o coñecemento do vocabulario en galego e castelán, á vez que entre os resultados acadados polos escolares galegos e os ofrecidos en castelán por Casanova e Rivera (1989) no estado español. A mostra estivo composta por 292 escolares de segundo curso no caso do castelán e 252 no caso do galego; 299 escolares de quinto nas probas de castelán e 267 nas de galego; 576 escolares de oitavo curso na valoración do castelán e 608 na do galego.

No estudo do coñecemento do vocabulario utilizou as mesmas fichas que Casanova e Rivera, traducidas no caso do galego. Esta tradución levouse a cabo aplicando os seguintes criterios: fíxose da forma máis literal posible; cando había posibilidade de usar máis dun término, sempre se usou o que se consideraba máis próximo ó nivel de idade para o que estaban destinadas as fichas; non se usaron palabras que non se consideraran compoñentes dun vocabulario básico galego. Non foi posible mante-la orde alfabética das palabras que tiñan na proba en castelán. En castelán aplicou 584 fichas en segundo curso, 598 en quinto e 576 en oitavo curso; en galego 504 en segundo, 533 en quinto e 608 en oitavo, de modo que en total foron aplicadas 3463 fichas.

O autor ofrece datos relativos á lingua galega e á castelán, entre outros aspectos, sobre: número de alumnos que responden en cada centro, número de fichas contestadas polos alumnos de cada centro, suma total de puntuacións ponderadas, puntuacións medias (resultado de dividi-lo número de fichas contestadas en cada centro polo número de alumnos que responde en cada centro), mediana de cada centro, desviación típica en cada centro, número de probas nulas, etc.

Non achega información directa sobre a estimación do tamaño do vocabulario, senón que máis ben céntrase na obtención de puntuacións brutas calculadas en tódolos cursos e ponderadas en 2.º e 5.º, en función das respostas correctas, e nas comparacións anteriormente mencionadas. Como fixeran Casanova e Rivera (Táboa 2.11.), presentan as puntuacións medias brutas e ponderadas, desdobrando estas puntuacións en 2.º e 5.º conforme non se tiveran en conta as fichas nulas (efectuando o cálculo en base ás puntuacións totais e ó número total de fichas respondidas) ou téndoas en conta (deducindo as fichas nulas do número total para efectua-lo cálculo).

Táboa 2.11. : Puntuacións medias brutas e ponderadas do vocabulario pasivo coñecido por escolares galegos, segundo Esteban (1997)

	Castelán				Galego			
	Punt. Bruta		Punt. Ponderada		Punt. Bruta		Punt. Ponderada	
	*	**	*	**	*	**	*	**
2º EXB	8,21	8,38	7,13	7,48	7,57	7,73	6,11	6,44
5º EXB	8,40	8,44	7,97	8,06	6,59	6,62	5,94	5,99
8º EXB	12,46				10,19			

* Puntuacións medias sen ter en conta as fichas nulas (cálculo da media en base ás puntuacións totais ó número de escolares avaliados).

** Puntuacións medias tendo en conta as fichas nulas (deducíndoas do número total para efectua-lo cálculo).

Conclúe o autor que o coñecemento do vocabulario en galego na poboación estudada é menor que o do castelán, así como que o coñecemento do vocabulario nesta lingua é menor que o que ten o alumnado da mostra estatal no estudo de Casanova e Rivera.

2.3.5 Suárez, Seisdedos e Meara (1998)

Estes investigadores elaboraron un test dos denominados de resposta Si/Non e achegan datos sobre o tamaño do vocabulario de alumnado entre 9 e 18 anos.

Utilizaron como mostra 822 escolares de quinto de EXB de Madrid e Galicia (inicialmente sobre uns 300 e 600 respectivamente) de tres niveis socioeconómicos (medio alto, medio e medio baixo e baixo) con idades entre 9-10 anos, o 90% dos mesmos, e 12 anos. Posteriormente utilizaron outras dúas mostras de 188 e 286 escolares que cursaban de segundo a cuarto de ESO (entre 13 e 18 anos).

Para a construción do test partiron do *Diccionario Escolar Etimológico* de García Hoz (1989) que contén 13.000 palabras do vocabulario usual (García Hoz, 1953), do vocabulario

escolar común en ensinanza primaria e media (García Hoz, 1977) e palabras de uso máis frecuente en español.

A continuación seleccionaron ó azar unha mostra do 10% destas palabras. Para elo, numeraron tódalas entradas do dicionario e de cada 100 elixiron ó azar 10 palabras utilizando unha táboa de números aleatorios. Deste modo, obtiveron unha mostra de 1954 palabras. O seguinte paso consistiu en pedirlle a tres xuíces (mestres con ampla experiencia profesional) que clasificaran cada palabra en catro categorías, en función da dificultade de aprendizaxe que consonte á súa opinión presentaran para o alumnado. Calculando as medias das puntuacións, na categoría A (maior facilidade) clasificáronse 533 palabras, na B 541, 620 na C e 260 na D (a de maior dificultade). A partir desta mostra elaboraron unha lista de 1000 pseudopalabras.

Posteriormente, utilizando unha táboa de números aleatorios, construíron 40 probas paralelas de vocabulario. Cada unha delas estaba composta por 40 palabras que representaban proporcionalmente ás palabras das catro categorías de dificultade e 20 pseudopalabras. Finalmente, tamén seleccionaron aleatoriamente 10 tests, que serían os que lle aplicaron á mostra de escolares.

Os resultados das estimacións aparecen na Táboa 2.12. Debe terse en conta que os datos que corresponden a 11-12 anos calculáronos mediante unha extrapolación dos obtidos polos grupos de idade de 9-10 a 13-14 anos. Como se pode observar, o aumento no número de palabras coñecidas sería dunhas 6400 palabras entre os 9-10 anos e os 18 anos, arredor de 1600 en cada grupo de idade, unhas 800 por ano.

Táboa 2.12.: Estimación do número de palabras segundo EVOCA

Idade	Estimación do número de palabras		
	Batería 1	Batería 2	Media
9-10 anos	6400	5600	6000
11-12 anos	7800	7800	7800
13-14 anos	9200	10400	9800
15-16 anos	10400	11200	10800
17-18 anos	12400	12400	12400

Capítulo 3

Tests de vocabulario e a súa utilidade

Introdución: enfoques e procedementos na avaliación do vocabulario

É necesario resaltar que existe unha relación estreita entre o contido que se expón neste capítulo e o que se comentou acerca de qué significa coñecer unha palabra (capítulo 1, punto 1.2.1). Esta relación baséase en que a concepción sobre qué significa coñecer unha palabra determina o tipo de procedemento que se utilice para a valoración do vocabulario, así como na propia construción do instrumento empregado para efectuar tal valoración. Tamén existe unha conexión entre o contido deste capítulo e o exposto no que se refire ós diferentes tipos de estudos sobre tamaño do vocabulario (capítulo 2, apartado 2.2). Esta conexión aséntase en que os tests que se comentan no presente capítulo, dalgún modo e mediante procedementos moi diferentes, teñen como finalidade estima-lo vocabulario, finalidade para a que tamén foron elaborados fundamentalmente os utilizados por García Hoz, Casanova e Rivera ou Suárez, Seisdedos e Meara, utilizando igualmente procedementos diferentes. Aquí abórdase o tema do coñecemento do vocabulario dende a perspectiva dos tests que se adoitan utilizar para esta finalidade, ofrecendo, con elo, unha panorámica xeral da cuestión, con obxecto de engarza-la contribución da segunda parte da tese neste corpo de coñecemento.

Existen dúas aproximacións principais no relativo á valoración do vocabulario, como xa se comentou no capítulo 2: tamaño e profundidade. Estas aproximacións adoitan utilizar instrumentos con características distintas. De tódolos modos, a gran maioría dos tests de vocabulario que existen tratan de efectuar unha estimación do seu tamaño.

As medidas cualitativas ou de profundidade do coñecemento de palabras teñen unha gran complexidade, en tanto que tratan de avaliar este coñecemento nas súas múltiples dimensións, en xeral, mediante escalas nas que a persoa avaliada debe indicar en que grao coñece cada unha desas dimensións dunha determinada palabra. A utilización deste tipo de medidas limita o número de palabras que é posible estudar nunha mesma proba á vez, o que vai en detrimento dunha visión máis ampla do dominio do vocabulario.

No que se refire ó enfoque que se centra no tamaño, utilízanse distintos procedementos, mediante os que se someten a comprobación un importante número de palabras. Estes procedementos inclúen diversas clases de tarefas e tipos de resposta, co obxecto de determina-lo número de palabras que coñece unha persoa ou en que lugar se encontra ese individuo respecto do grupo da súa idade en función da puntuación alcanzada.

Un dos procedementos é o utilizado nos *tests de elección múltiple*. A tarefa en cada elemento destes tests consiste en elixir entre un grupo de palabras a que representa o significado doutra palabra determinada ou en completar frases con palabras dadas. O formato permite varia-la dificultade dos ítems manipulando os *distractores* (alternativas incorrectas incluídas dentro dun determinado grupo de elementos entre os que se debe de escoller). Estas probas presentan fundamentalmente dous inconvenientes. O primeiro é que non resultan moi fiables no relativo ós ítems, posto que é posible, por exemplo, que unha persoa que coñeza unha palabra que se lle presenta conteste erroneamente porque non entende o resto das palabras ou porque non coñece o significado preciso requirido nese ítem concreto. Cando a tarefa consiste en completar frases, é posible contestar correctamente por exclusión, sen coñecer realmente a palabra adecuada, se se coñecen o resto das palabras. O segundo inconveniente está relacionado coa dificultade para efectuar estimacións do vocabulario total que poda coñecer unha persoa, basicamente porque a proporción de palabras que se poden utilizar nunha proba destas características é limitada.

Outro tipo de procedemento é o representado nos *tests de fluidez verbal* (*spew tests*). A tarefa consiste en que o suxeito emita nun período determinado de tempo tódalas palabras que sexa capaz e que empecen por unha letra ou que pertencen a unha determinada categoría. Se ben, en principio, cabe esperar que puidera servir de base para determina-lo vocabulario dunha persoa, as correlacións obtidas coas medidas realizadas con outro tipo de tests son moi baixas (English, 1985). Alén disto, presentan problemas de índole práctica e metodolóxica, como son que o número de palabras que se produzan está moi afectado por variables como o esforzo e motivación persoais, non existe un número equivalente de palabras que teñen como inicial cada unha das letras do alfabeto e debería de estar previamente determinado que tipo de palabras se contabilizan como distintas (derivadas, sinónimos, formas dialectais, etc.).

Un terceiro tipo de procedemento é o que utiliza unha tarefa de recoñecemento de palabras. Consiste en decidir se entre unha determinada serie de estímulos hai palabras que se coñecen ou non. Adoptando como base este tipo de tarefas, desenvolvéronse os denominados *tests de resposta Si/Non*, ós que se fará referencia con maior profundidade e amplitude posteriormente.

Existen, así mesmo, outros procedementos como o que se segue en probas nas que a tarefa consiste ben en relacionar sinónimos ou palabras que gardan unha conexión semántica, ben en dici-lo significado dunha palabra estímulo mediante unha resposta aberta, ben en emiti-lo nome do que se representa en debuxos ou ben en sinala-lo debuxo unha vez que o investigador dí o nome.

Entre todo este tipo de probas é posible establecer unha clasificación atendendo ó que se pode considerar a finalidade para a que se utiliza a información que permiten obter sobre o vocabulario. Deseguido preséntase unha proposta de clasificación atendendo a este criterio e expónse sucintamente o contido dalgúns tests que se poden encadrar en cada categoría, ademais dalgúns datos sobre a súa elaboración.

3.1 Tests de tamaño relativo

A pretensión principal destes tests é obter puntuacións que permitan comparalas conseguidas por un suxeito coas puntuacións do grupo normativo de idade correspondente, máis que obter estimacións do tamaño absoluto do vocabulario que coñecen os individuos.

3.1.1 Tests de intelixencia xeral e aptitudes

Inclúense neste grupo os tests de vocabulario que están integrados noutros tests, os cales teñen como finalidade primordial a medición da intelixencia e aptitudes. A continuación descríbense brevemente algúns deles.

1. *Test de Pronóstico Académico APT*, de Bennet, Clendenen, Doppelt, Ricks, Seashore e Wesman (1985). A proba de "Factor Verbal" consta de 60 ítems, nos que se solicitan diferentes tipos de asociacións mediante sinónimos, contrarios e relacións por analoxías, debendo ser elixida unha palabra entre catro dadas (*Madrid es a provincia como Galicia es a ...; Pájaro es a pez como avión es a ...; Singular es a plural como individuo es a ...*).

2. *Test de Aptitudes Escolares TEA*, de Seisdedos, de la Cruz, Cordero e González (1990). Contén unha proba de "Vocabulario" (sinónimos), que consta de 22 ítems de elección múltiple no Nivel 1 e de 30 ítems non Niveis 2 e 3, nos que é preciso seleccionar entre un grupo de catro a palabra sinónima da que se propón como estímulo. A proba denominada "Palabra Diferente" do mesmo test consta de entre 18 e 20 ítems, cada un dos cales está formado por un grupo de 4 ou 5 palabras conforme ó nivel. A tarefa do suxeito examinado consiste en sinala-la palabra que sexa diferente atendendo ó concepto ou significado. Non é estritamente unha proba de vocabulario, pero si se pode considerar que

en certo modo avalía competencia en vocabulario, na medida en que trata de comproba-lo coñecemento sobre as relacións en canto ó significado entre palabras.

3. *Test de Aptitudes Psicolingüísticas de Illinois (ITPA)*, de Kirk, McCarthy e Kirk (1989). Forma parte deste test unha proba específica de vocabulario, que os autores denominan Proba de "Fluidez Verbal". A tarefa consiste en dicir tódalas palabras que o neno ou nena sexa capaz, primeiro libremente e logo atendendo a tres categorías (partes do corpo, animais e froitas).

Este test contén así mesmo outras dúas probas moi relacionadas co coñecemento do vocabulario. Unha é a proba de "Asociación Auditiva", que consta de 40 ítems nos que a tarefa consiste en completar frases mediante analoxías (*El pan se come, la leche...*, por exemplo). A outra proba é a denominada proba de "Integración Gramatical", constituída por frases incompletas que se deben completar (*Donde se venden zapatos es una zapatería, donde se vende pescado es una...*, por exemplo).

4. *Escala de Inteligencia de Wechsler para Niños (WISC)*, (Wechsler, 1999). Contén diferentes probas nas que o peso do vocabulario pode ser considerado importante, como son "Información", "Semellanzas" e "Comprensión", en tanto que para resolvelas é preciso dispoñer de termos asociados e de coñecemento sobre as cuestións abordadas, pero hai unha específica de vocabulario, denominada precisamente así, "Vocabulario". Esta proba consta de 20 ítems formulados de modo que se debe defini-lo termo que se presenta (*¿Qué es un cuchillo?, ¿Qué significa valiente?*, por exemplo).

5. *Escala McCarthy de Aptitudes e Psicomotricidade para Nenos (MSCA)*, de McCarthy (1996). Entre as probas que a compoñen hai varias denominadas de vocabulario, pero contén, amais, algunha outra moi relacionada co coñecemento do mesmo.

- "Proba de Vocabulario". A primeira parte denomínase "Vocabulario Pictórico". Consta de cinco láminas e a tarefa consiste en sinalar ou responder denominando os debuxos correspondentes. A segunda parte denominada "Vocabulario Oral", é unha proba na que a persoa examinada debe definir 10 termos ante a pregunta de *¿Qué es...? ou ¿Qué quiere decir...?*

- "Proba de Fluencia Verbal", na que a tarefa consiste en dicir tódalas palabras posibles, atendendo a catro categorías diferentes (alimentos, animais, prendas e vehículos).

- "Proba de Opostos", onde se trata de completar oralmente nove frases utilizando antónimos.

- "Proba de Conceptos", que, dalgún modo, esixe un coñecemento do vocabulario comprensivo relacionado con cores e tamaños, xa que consiste en sinalar figuras xeométricas segundo o tamaño e a cor a medida que se van solicitando.

6. *Test de Factor G. Escala 1*, de Cattell e Cattell (1989). Neste test hai unha proba de "Identificación de Debuxos", formada por 10 ítems, que se teñen que sinalar unha vez que o examinador os denomina, e outra de "Adiviñas" (12 ítems), na que se ofrecen tres palabras e a continuación fórmase unha pregunta que se reponde cunha das palabras dadas (por exemplo, *1. Hierbas, bombillas, hojas. ¿Cuáles de estas cosas son verdes y crecen en los árboles?*).

Existen por outra parte instrumentos como o *Test Boehm de Conceptos Básicos* (1988), que, aínda que non é estritamente de vocabulario, pódese considerar que efectúa unha avaliación do mesmo, se ben limitándose a determinados conceptos espaciais, temporais e de cantidade, fundamentalmente.

3.1.2 Tests específicos de vocabulario

Nesta categoría inclúense aqueles que teñen como propósito central a avaliación do vocabulario. Entre eles, coméntanse a continuación algúns de forma sucinta.

1. *Test de Vocabulario*, de García Hoz (1946), ó que xa se fixo referencia anteriormente (capítulo 2, punto 2.3.1).

Consta de 100 ítems, entre os que predominan os de elección múltiple. Destes 100 ítems, os 20 primeiros consisten en emparellar unha palabra estímulo cunha das opcións que se ofrecen nun grupo de cinco, atendendo a que exprese unha idea semellante á primeira, co que, en certo modo, o que esixe é identificar sinónimos. Os seguintes 20 ítems adoptan o formato dunha definición incompleta que se debe completar cunha das cinco palabras que aparecen como alternativa. A terceira parte consta de 4 ítems formados por un grupo de 7 palabras e outro grupo de 5 palabras; a tarefa consiste en escribir ó lado de cada unha das palabras do segundo grupo as que máis se parezan do primeiro grupo. A cuarta parte da proba está composta por 20 ítems tipo *cloze*, é dicir, aparecen frases incompletas e a tarefa esixe que se seleccionen palabras dun grupo de cinco para completa-la frase. A última parte está formada por 20 ítems e a tarefa é idéntica que para os 20 primeiros ítems, unicamente varían as palabras, é dicir, trátase de seleccionar entre un grupo de 5 palabras a “que mellor diga o que significa” unha palabra estímulo que se ofrece inicialmente.

O autor achega datos que representan as puntuacións medias logradas polos distintos grupos diferenciados por sexo, idade (de 9 a 18 anos) e nivel socioeconómico, o que, dalgún modo, pode utilizarse coma un baremo moi elemental. Pero, tamén ofrece a posibilidade de efectuar unha estimación do número de palabras coñecidas por un determinado escolar ou grupo, multiplicando a puntuación obtida por 48.300 (as palabras do dicionario da RAE de 1939) e dividindo por 100.

2. *Proba de Vocabulario*, de Pérez González (1981). Forma parte dunha batería de probas escolares para 8.º de EXB (comprensión lectora, exactitude lectora, vocabulario, ortografía, cálculo, aplicacións matemáticas e unha escala de hábitos e técnicas de estudo). Aplicóuselle na súa totalidade a 3096 escolares de 8 provincias (Alacante, Cádiz, Cuenca, Xaén, León, Madrid, Soria e Valencia).

Foi elaborada a partir de 575 vocábulos seleccionados de textos de Lingua Castelán para 6.º, 7.º e 8.º de EXB de catro editoriais, mediante un procedemento ou sistema de xuíces. Desta mostra seleccionaron ó azar un 10%, de modo que, por redondeo, quedaron 60 vocábulos. Ademais, seleccionáronse outras 20 do *Vocabulario Común* de García Hoz. En total, a proba está composta por 80 ítems. Cada un deles consta dunha palabra estímulo e outras catro entre as que se debe seleccionar a que signifique o mesmo que o estímulo. Ofrece puntuacións medias totais e por grupos (sexo, titularidade do centro, tipo de centro polo tamaño).

3. *Test de Vocabulario en Imaxes de Peabody* (Dunn, 1985). Trata de medi-lo vocabulario de suxeitos de entre 2 anos e 6 meses e 18 anos. Foi elaborado inicialmente en inglés, seleccionando 3885 palabras do *New Collegiate Dictionary* de Webster, ás que se lle engadiron outras 1200 palabras extraídas de listas de palabras e de estudos sobre coñecemento do vocabulario de nenos e mozos. Do total de palabras, o autor extraeu 684 e construíu un test de elección múltiple, que consiste nunha lámina con catro imaxes para cada palabra. As palabras estímulo abarcan 18 categorías, incluíndo, por exemplo, nomes de obxectos, situacións, profesións e animais, accións e atributos, transportes, etc. A tarefa consiste en sinala-la imaxe que se corresponde coa palabra estímulo presentada oralmente.

Para a versión española seleccionáronse 150 láminas e aplicóuselle o test a unha mostra de 356 escolares madrileños. A fiabilidade foi calculada mediante o procedemento de dúas metades (elementos pares e impares) e o procedemento test/retest. Os coeficientes de

fiabilidade obtidos foron, respectivamente, 0,96 e 0,77. A validez de contido susténtase no proceso de construción; a de construto baséase nunha evidencia indirecta: as puntuacións incrementáanse coa idade; a concurrente foi calculada mediante a correlación con outros tests de vocabulario e con tests de intelixencia: o valor medio das correlacións con tests de vocabulario foi de 0,71 e con tests de intelixencia, como son o WISC e a Escala Stanford-Binet, foi de 0,60 e de 0,72, respectivamente.

Inclúe baremos en centís e en puntuacións típicas, desde os 2 anos e 6 meses, con información para períodos de dous meses en dous meses, ata os 18 anos e 11 meses, con información de seis meses en seis meses a partir dos 15 anos.

4. *Proba de Casanova e Rivera* (1989). Xa se presentou no capítulo 2, punto 2.3.3. Aquí será suficiente con comenta-la de forma resumida. Para a valoración do vocabulario activo utilizaron dúas probas: unha proba de escritura espontánea, na que a tarefa consiste en que os suxeitos escriban durante 5 ou 10 minutos (segundo sexan do 1.º ciclo ou dos dous seguintes ciclos de EXB) tódalas palabras que poidan sobre calquera tema que desexen; unha segunda proba de escritura suxerida, na que a demanda é a mesma que na anterior proba, pero centrándose nun tema concreto, entre varios que se propoñen. A proba para a valoración do vocabulario pasivo está formada por 35 fichas con catro ítems en cada unha. En 2.º e 5.º de EXB, en cada ficha valóranse 10 palabras do vocabulario elaborado polas propias autoras (unha palabra no primeiro e segundo ítems, tres palabras no terceiro e cinco palabras no cuarto ítem). A tarefa consiste en que expliquen doutra forma unha frase, explica-lo significado dunha palabra subliñada nunha frase, completar unha frase (en 2.º) ou explicar 3 palabras (en 5.º), e relacionar palabras con frases polo significado. No caso do alumnado de 8.º de EXB, cada ficha contén 15 palabras (tres no primeiro e segundo ítems, catro no terceiro e cinco palabras no cuarto ítem). A tarefa consiste en que os suxeitos

expliquen o significado de palabras subliñadas en frases, dicir sinónimos de palabras dadas, explica-lo significado de palabras e relacionar palabras con frases polo significado.

Proporcionan puntuacións medias por grupos, polo que cabe a posibilidade de compara-la situación dos suxeitos co grupo de idade que lle corresponde.

5. *Escala de Vocabulario de Pastora* (1990). Consta de cinco probas compostas por tarefas que inciden no que o autor denomina diferentes campos (sinonimia, antonimia, composición, derivación, polisemia e homonimia). As palabras utilizadas en cada unha das probas foron seleccionadas por un equipo de docentes con cinco anos de experiencia, cando menos.

Polo que respecta á proba de sinonimia, consta de 50 ítems. En cada un deles hai un vocábulo estímulo e un grupo de tres máis entre os que se debe elixir o que ten un significado equiparable ó estímulo. A de antonimia consta tamén de 50 ítems, consistentes nun vocábulo estímulo e outros tres para selecciona-lo que significa o contrario que o estímulo. A composición valórase cunha proba que consta de dúas partes: na primeira aparecen dez parellas de palabras para relacionalas formando palabras compostas; na segunda simplemente se pide que formen tódalas palabras compostas que lles sexa posible. Para a derivación utilizouse unha proba formada por 45 vocábulos, que son palabras compostas e a tarefa consiste en escribir que palabras simples as compoñen, alén de escribi-lo seu significado. Finalmente, a proba de polisemia e homonimia consiste en escribir frases con 25 palabras dadas, de forma que expresen tantos significados diferentes da palabra de referencia como sexa posible.

O autor ofrece as puntuacións medias que lle corresponderían ó alumnado da illa de Gran Canaria en cada proba, que, ó seu xuízo, é posible xeneralizar á comunidade de Canarias, se ben no texto do que se extraeu a información non se explicitan as idades. Pastora tamén presenta o que denomina baremos, onde figuran puntuacións medias (non

aparece claramente explicado como se obtiveron, agás a media xeral, que fai corresponder coa cualificación académica de *Suficiente*) e a situación destas puntuacións medias nunha escala por tramos que se corresponde coas cualificacións oficiais desde *Moi deficiente* ata *Sobresaliente*.

3.2 Tests de estimación do tamaño absoluto

3.2.1 Tests de García Hoz (1946, 1976)

Como se sinalou no punto 3.1.2, o *Test de Vocabulario* de García Hoz de 1946 tamén permite efectuar unha estimación do número de palabras coñecidas por unha persoa. O *Test de Vocabulario Usual* (1976) está construído sobre a base do *Vocabulario usual* do mesmo autor e formado por unhas 13.000 palabras. Para a elaboración deste último test, en primeiro lugar, dividiu o vocabulario usual en catro partes, segundo a frecuencia das palabras; a continuación seleccionou as palabras, formando doce series de 100 vocábulos cada unha e, finalmente, utilizou unicamente unha destas series na construción do test.

A estrutura é idéntica á do *Test de Vocabulario* de 1946, variando as palabras utilizadas. Permite calcula-lo número de palabras coñecidas do vocabulario usual. Tendo en conta que cada unha das palabras da proba representa 130 do vocabulario usual, para realiza-lo cálculo, multiplícase a puntuación obtida na proba por 130.

3.2.2 Test de Diack (1975)

É dos primeiros exemplos dun dos métodos máis directos de medi-lo coñecemento de palabras, xa que a persoa avaliada unicamente debe responder se coñece ou non unha determinada palabra.

Para a construción dividiu en seis niveis as 36.000 palabras do *Diccionario Conciso de Oxford*, de modo que puidese ser probable que tódalas palabras do primeiro nivel fosen

adquiridas antes que as do segundo e así sucesivamente. Para verificar que ocorría así, levou a cabo comprobacións no Reino Unido con alumnado que cursaba desde primaria a estudos universitarios. Consta de 50 subtests de 60 palabras cada un, divididos nos seis niveis de dificultade mencionados (10 palabras por nivel, ordenadas de menor a maior dificultade). Cada nivel representa 6000 palabras, mentres que cada ítem representa 600. Na Figura 3.1. ofrécese unha exemplificación das palabras incluídas nos niveis 1 e 6 do test número 8.

Nivel 1	Nivel 6
1. apology	51. calamite
2. cartoon	52. dugong
3. eagerly	53. ecdysis
4. message	54. heriot
5. naturalist	55. meuse
6. parachute	56. olibanum
7. preparation	57. pomace
8. quarry	58. sclerotic
9. target	59. tylopodus
10. twilight	60. uprhoe

Figura 3.1.: Palabras incluídas nos niveis 1 e 6 do test número 8 de Diack (1975)

É interesante non só polo procedemento que utiliza na avaliación senón tamén polo procedemento seguido para a mostraxe das palabras, o que permite realizar estimacións relativamente razoables do tamaño do vocabulario. Segundo o autor, os adultos con educación secundaria situaríanse no nivel 3 (12.000-18.000 palabras), mentres que o nivel 4 (18.000-24.000 palabras) require unha educación universitaria, o nivel 5 (24.000-30.000) alcázanlo as persoas en escalas profesionais superiores e o nivel 6 moi poucas persoas. Non

obstante, a xuízo de Goulden, Nation e Read (1990) estes cálculos, se ben non son moi desacertados, tampouco son moi axustados, en tanto que nin levou a cabo unha mostraxe das palabras coas necesarias precaucións, nin tivo en conta as palabras base (aquelas utilizadas como entradas principais nos dicionarios ou as formas menos flexionadas dun grupo de palabras relacionadas).

O procedemento é de autoavaliación e comézase polo nivel 1. A tarefa consiste en ir sinalando as palabras que **Non** se coñecen ata sinalar 10 seguidas, momento en que se debe parar e volver ás 5 últimas sinaladas como coñecidas e definilas mediante sinónimos, usándoas nunha frase, facendo un diagrama ou un debuxo. Se contrastando cun dicionario se comproba que todas ou a maioría delas son correctas, asúmese que as sinaladas anteriormente como coñecidas tamén o son. A puntuación obtense contabilizando o número total de palabras coñecidas, subtraéndolle as que están equivocadas. Para mellora-la fiabilidade, recoméndase utilizar ó menos tres tests e calcula-la puntuación media para estimalo tamaño do vocabulario, multiplicando a media por 600.

3.2.3 Tests de resposta Si/Non

A tarefa ten certa semellanza coas de decisión léxica, aínda que o que requiren decidir sexa de diferente natureza. Os tests están compostos por listas de ítems, unha porcentaxe dos cales son palabras reais e outra porcentaxe son pseudopalabras. A tarefa dos suxeitos consiste en sinala-las palabras que están seguros de coñecer. Se un suxeito sinala unha palabra como coñecida, pódese supoñer que, polo menos, coñecerá que é da súa lingua, o que representa o coñecemento mínimo que se pode ter da palabra. Este coñecemento está na base e del dependen outros niveis superiores de coñecemento. Polo tanto, estes tests tratan de comproba-lo nivel máis básico da competencia en vocabulario (Meara, 1994), a competencia máis esencial da que dependen o resto das competencias. Se, por exemplo, unha persoa non é capaz de recoñecer que *casa* é unha palabra en galego,

difícilmente poderá realizar calquera outra tarefa con ela que sirva para comprobar en que medida a coñece. Neste tipo de probas non se parte dunha concepción estrita acerca da clase de coñecemento requirido para considerar que un suxeito coñece unha palabra de determinada lingua. Deste modo, cando, polo menos, se recoñece que unha forma é realmente unha palabra, xa representa un certo coñecemento da mesma, polo que se podería admitir como coñecida. Isto concorda coa idea de qué é unha palabra mantida por Bloomfield, que a entende como a “mínima forma libre”, dándolle maior peso ó aspecto formal que ó significado na definición desta unidade lingüística. Por outra parte, seguindo a semioloxía integracionista, é conveniente recordar tamén que os humanos construímos os significados negociándoos nos usos e situacións concretas, polo que, en boa medida, importa o significado que cada quen lle atribúe (Harris, 2000). É de interese, ó meu xuízo, o que Harris expresa ó respecto:

A base do pensamento tradicional é ó que o integracionismo se refire como a falacia do código fixo. Esta é a noción de que a comunicación (en particular a comunicación lingüística) depende do establecemento dun sistema de correlacións entre formas e significados recoñecido publicamente. Así, é asumido que a lingua (inglés, francés, suahili, etc.) é un código fixo que, relacionando entidades nunha dimensión chamada ‘forma’ con entidades nunha dimensión chamada ‘significado’, fornece ós usuarios da lingua de medios de transmisión e recepción de ideas (p. 72).

A base dunha semioloxía integracionista é o recoñecemento de que toda comunicación é intrinsecamente dependente do contexto. O signo é en si mesmo un produto do contexto (...). Os signos, vistos desde esta perspectiva, non son asociacións de forma e significado como un código preestablecido. Os signos afloran de situacións e circunstancias reais nas que os participantes están inmersos. Así, nunha situación concreta, calquera obxecto, acción, etc. pode adquirir valor semiolóxico. É máis, non é esencial que tódolos participantes estean de

acordo no seu valor, ou, incluso, en que as características dunha situación teñen determinado valor (2000, p. 81).

Deseguido revísanse con maior amplitude a partir do exposto por Vives (1995) algúns dos tests existentes deste tipo, coa finalidade de explicar de forma máis pormenorizada, non só en que consisten, senón tamén a evolución en canto ó contido e proceso de construción.

1. Zimmerman, Broder, Shaughnessy e Underwood (1977) intentaron demostrar a fiabilidade dun test de recoñecemento de vocabulario en lingua materna que non esixía respostas dicotómicas e baseado en “Medidas de Detección de Sinais”.

O test constaba de 300 palabras e 100 pseudopalabras e foi aplicado a 200 suxeitos universitarios, que debían sinalar se recoñecían ou non as palabras, utilizando unha escala de seis niveis para expresa-lo grao de certeza no xuízo (Figura 3.2.).

1. Estás seguro de que non é unha palabra.
2. Pensas que non é unha palabra.
3. Supos que é probable que non é unha palabra.
4. Supos que probablemente é unha palabra.
5. Pensas que é unha palabra.
6. Estás seguro que é unha palabra.

Figura 3.2.: Escala de coñecemento dunha palabra (Zimmerman et al., 1977)

Ademais, recolleron as puntuacións do *Scholastic Aptitude Test (SAT)* e as cualificacións que obtiveran no ámbito verbal ó remate de Bacharelato (*High School*). Na obtención de puntuacións contabilizaron os acertos (sinalar con 4, 5 ou 6 as palabras e con 1, 2 ou 3 as pseudopalabras) e os erros ou falsas alarmas (o sinalar pseudopalabras con 4, 5

ou 6 e con 1, 2 ou 3 as palabras). A combinación das puntuacións de acertos e erros permitiulles trazar unha curva de rendemento para cada suxeito que reflectía a súa tendencia de resposta. Ó respecto, detectaron unha tendencia a responder predominantemente utilizando números altos ou baixos da escala, debido seguramente a que os suxeitos tiñan dificultades para manter na memoria toda a escala e efectuar de forma simultánea xuízos consistentes. Obtiveron unha medida alternativa: primeiro calcularon a media dos xuízos dados ás palabras e ás pseudopalabras; despois calcularon as diferenzas entre as medias das dúas distribucións e estandarizáronas. Esta medida resultou ser independente da tendencia de resposta dos suxeitos e da variabilidade nos xuízos. A esta diferenza estandarizada denominárona $\delta\sigma$, indicador da capacidade de discriminación entre os dous tipos de estímulos: se a media de palabras fose 6 ou próxima e a de pseudopalabras 1 ou próxima, consideraríase a ese suxeito como moi seguro nas súas respostas. Dito doutro modo, unha diferenza grande entre estas dúas medias indicaría que quen reponde discrimina ben entre os dous tipos de estímulos. A correlación entre a curva de rendemento de cada suxeito e as diferenzas estandarizadas das súas puntuacións foi de 0,96.

Por outro lado, tamén obtiveron unha correlación de 0,48 entre as diferenzas estandarizadas e os resultados no test *SAT*. Zimmerman et al. atribúen esta moderada correlación a que o seu test mide vocabulario, mentres que o test *SAT* é unha medida máis xeral da linguaxe e outras aptitudes. Utilizando a escala como unha proba de resposta dicotómica (asumindo que os ítems sinalados con 4, 5 e 6 eran considerados como palabras e que os sinalados con 1, 2 e 3 eran considerados como pseudopalabras), as correlacións foron máis baixas, polo que concluíron que a proba de resposta dicotómica non proporcionaba datos de tanta calidade como os obtidos coa escala. Agora ben, non ofrecen información concreta de “canto” menor foi a correlación coa forma dicotómica. Amais, a forma dicotómica non é de resposta Si/Non, senón unha reformulación da escala, co que a asunción de que as categorías máis altas correspóndense a *palabra* e as máis baixas a

pseudopalabra non parece moi clara, máis cando aparece a tendencia xa comentada na resposta dos suxeitos a sinalar números altos ou baixos da devandita escala.

A pesar de certas dificultades que se poden presentar en canto á efectividade e á consistencia nas medicións realizadas cunha escala graduada, o estudo introduce por primeira vez a aplicación dos principios da “Teoría de Detección de Sinais” en tests de recoñecemento de vocabulario, abrindo unha nova liña de investigación que tivo diferentes seguidores como Anderson e Freebody ou Meara, aínda que estes usan tests de resposta dicotómica Si/Non.

2. Anderson e Freebody (1983) elaboraron unha proba para estima-lo tamaño do vocabulario en lingua materna. A súa proba representou un punto de partida dentro dunha nova concepción dos tests de tamaño de vocabulario. Proporcionou, así mesmo, unha base sólida para desenvolver posteriormente a técnica no ámbito da avaliación de aprendices dunha segunda lingua.

Os autores consideran que o principal problema na avaliación do tamaño do vocabulario é a diferenza no número de palabras de que constan os dicionarios empregados para selecciona-los ítems a incluír nos tests, xunto coa dificultade de definir operativamente que é unha palabra, xa que os corpus poden variar en miles de palabras. Isto ten repercusión no número de palabras que se utilicen porque se, por exemplo, se parte dun corpus de 20.000 palabras, unha mostra do 1% serían 200 palabras, mentres que se se parte dun corpus de 50.000, o 1% serían 500 palabras. Un test deste tamaño non resultaría operativo, a non ser que se reducise o número de ítems, pero, deste modo, a mostra de palabras non sería tan representativa. Se as mostras de palabras varían moito dunha investigación a outra, as estimacións da proporción de palabras coñecidas tamén variarán.

Anderson e Freebody propoñen a utilización dun test de resposta Si/Non, que consiste nunha lista de control na que os suxeitos deben sinala-las palabras que coñecen,

incluíndo pseudopalabras, como medida de control da sinceridade. A corrección proporciona dúas medidas: *hits* (palabras reais que sinalan coma coñecidas) e *falsas alarmas* (pseudopalabras que sinalan coma coñecidas). Posteriormente, realízase un cálculo mediante a fórmula que segue, obtendo a proporción de palabras realmente coñecidas:

$$P(K)=[P(H) - P(FA)] / [1 - P(FA)]$$

onde $P(K)$ representa a proporción de palabras coñecidas, $P(H)$ a proporción de acertos e $P(FA)$ a proporción de falsas alarmas.

En comparación cos tests de elección múltiple, estes tests presentan diversas vantaxes: son relativamente fáciles de construír, aplicar, corrixir e puntuar; requiren pouco tempo para a aplicación; permiten utilizar un maior número de palabras que outro tipo de probas, o que posibilita a comprobación de máis palabras no mesmo tempo; minimizan as esixencias en canto a coñecemento estratéxico ou de habilidades de expresión.

Compararon as puntuacións obtidas por unha mostra de 120 escolares de 5.º de ensino primario nun test de elección múltiple e nun test de resposta Si/Non, con obxecto de comprobar cal estimaría mellor o tamaño do vocabulario. A correlación foi de 0,84, que non é moi alta, tendo en conta que utilizaron as mesmas 195 palabras para a construción dos dous tests.

Para calcula-la validez, compararon as puntuacións obtidas en base á resposta a 40 palabras de cada proba coas puntuacións obtidas mediante unha entrevista na que se lle pedía que definiran as palabras, que as usaran en frases ou dixeran algo sobre elas. Puntuaron as respostas na entrevista de acordo con tres criterios (estricto, moderado e brando). As correlacións entre as respostas acertadas no test tipo Si/Non e as puntuacións na entrevista son claramente máis altas que as correlacións entre as puntuacións do test de elección múltiple e as da entrevista, o que reflectiría que os test tipo Si/Non son mellores indicadores do número de palabras realmente coñecidas por un suxeito. Non obstante, segundo os autores, presentan unha serie de limitacións, como son que non distinguen entre

diferentes significados dunha palabra, valoran esencialmente vocabulario pasivo e non son totalmente adecuados para avalia-los efectos de instrución directa (pódense recoñecer as palabras pola familiaridade da forma máis que polo coñecemento do significado).

Os autores analizaron as *falsas alarmas* e encontraron patróns de resposta distintos en función do nivel de competencia en lectura. Con todo, a gran maioría dos suxeitos, pero especialmente os bos lectores, seleccionaron principalmente as pseudopalabras que morfoloxicamente se asemellaban máis a palabras derivadas reais que a outro tipo de pseudopalabras. O feito de que se sinalen nun número elevado os estímulos que máis se asemellaban morfoloxicamente a palabras derivadas reais mostra que os suxeitos intentan recoñecer-lo estímulo en base ós compoñentes morfolóxicos, identificándoos como pertencentes a categorías gramaticais particulares: nome, verbo, etc.

Os autores conclúen que a puntuación obtida nos tests tipo Si/Non son un mellor indicador do coñecemento das palabras sometidas a comprobación que os tests de elección múltiple, á vez que permiten efectuar estimacións máis precisas do número de palabras coñecidas. Alén disto, a simplicidade e rapidez para a súa administración convértenos nunha útil solución ó problema de cómo comproba-lo coñecemento do vocabulario en lingua materna con mostras amplas de palabras. A pesar das limitacións, funcionan como instrumentos robustos e fiables e permiten sentar as bases para unha prometedora liña de investigación.

3. White, Wayne e Graves (1989) levaron a cabo un estudo con 278 suxeitos de diferente idade e nivel de escolarización (1.º, 2.º, 3.º e 4.º), procedencia social e habilidade lectora, coa finalidade de validar un test do tipo Si/Non para medir o vocabulario en lingua materna. Para elo, basicamente, compararon os resultados no test de resposta Si/Non cos obtidos nunha proba de elección múltiple.

Utilizaron 56 palabras extraídas das 19.500 palabras máis frecuentes do *American Heritage Word Frequency Book* (Carrol, Davies e Richman, 1971). Estratificaron o corpus en 7 bloques que se incrementaban en tamaño (150, 300, 600, 1200, 2400, 4800 e 9600) e seleccionaron 8 palabras ó azar de cada bloque. Amais, utilizaron 44 pseudopalabras das usadas por Anderson e Freebody (1983). Os 56 ítems da proba de elección múltiple consistían en frases cun espazo en branco e catro palabras alternativas entre as que elixir para completa-la frase.

Empregaron unha entrevista para comproba-lo coñecemento das palabras por parte dos suxeitos. A entrevista estaba dividida en dúas partes, cada unha das cales contiña a metade das palabras do test (28), á vez que estaban representados en ámbalas dúas partes os diferentes niveis de frecuencia das devanditas palabras. A tarefa consistía en definir palabras representadas en fichas e usalas nunha frase, seguido dunha pregunta, co que se requiría unha pronunciación correcta e algún coñecemento do significado da palabra.

Usaron tres medidas dependentes: puntuacións do test Si/Non, puntuacións da proba de elección múltiple e os datos obtidos na entrevista (analizando só 28 palabras). As puntuacións do test tipo Si/Non foron calculadas mediante a mesma fórmula que utilizaron Anderson e Freebody (1983): $P(K) = [P(H) - P(FA)] / [1 - P(FA)]$.

Na puntuación do test de elección múltiple corrixiron a sinceridade coa fórmula $C - [E / (n - 1)]$ (Downie e Health, 1983), onde C representa o número de respostas correctas, E as respostas erróneas e n o número de alternativas de resposta.

As medias obtidas co test de resposta Si/Non e coa proba de elección múltiple son maiores que as obtidas coa entrevista. As obtidas co test Si/Non son máis próximas ás acadadas na entrevista que as obtidas coa proba de elección múltiple. Os resultados no test Si/Non son tamén máis consistentes nas diferentes mostras e a sobreestimación non varía significativamente, aínda que, en principio, tanto este test coma a proba de elección múltiple producen unha sobreestimación en relación coa entrevista.

Atendendo a todos estes datos, os autores conclúen que a validez do test é alta, resultando ser, en términos xerais, un indicador máis axustado que os proporcionados pola proba de elección múltiple, e máis constante en diferentes mostras de suxeitos.

4. Meara e Buxton (1987) utilizaron no Reino Unido un test de reposta Si/Non para medi-lo vocabulario nunha segunda lingua (inglés) de 100 estudantes, cunha media de idade de 22 anos. Pódese considerar como o primeiro intento de calcula-lo tamaño total do vocabulario nunha segunda lingua. Utilizaron unha proba de elección múltiple de 25 ítems e un test de tipo Si/Non de 100 ítems (60 palabras reais e 40 pseudopalabras) e compararon os resultados cos obtidos no *Cambridge First Certificate in English*. A correlación entre o test tipo Si/Non e a proba de elección múltiple foi 0,703, o que interpretaron coma que os dous test miden a mesma clase de habilidades. O feito de que fose máis baixa que a obtida por Anderson e Freebody (1983), que foi de 0,84, atribúenllo á falta de homoxeneidade do grupo e a que os dous test difiren nas palabras utilizadas (cun grupo máis amplo de suxeitos a correlación que obtiveron foi de 0,829). Da comparación coas puntuación en *Cambridge First Certificate* conclúen que o test Si/Non é máis útil para discriminar entre candidatos que a proba de elección múltiple e que resulta tamén un instrumento con maior capacidade para predici-las puntuacións no *Cambridge First Certificate* (Figuras 3.3. e 3.4.).

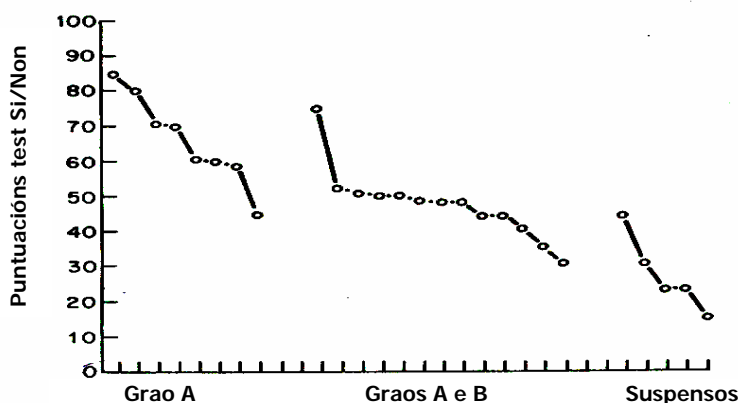


Figura 3.3.: Comparación entre as puntuacións en *First Certificate* e nun test Si/Non

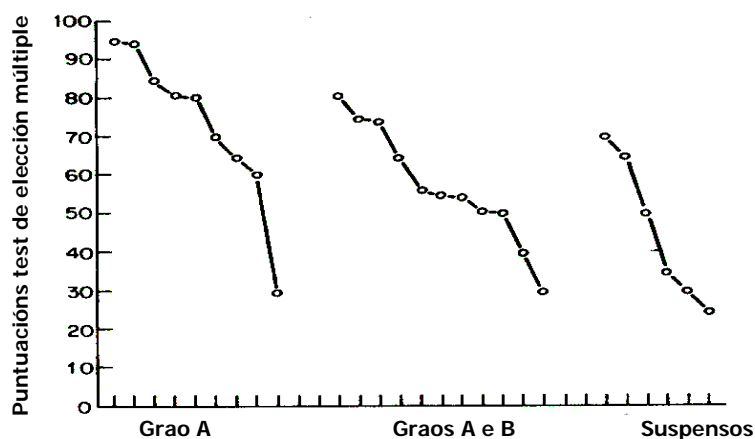


Figura 3.4.: Comparación entre as puntuacións en *First Certificate* e nun test de elección múltiple

Os tres últimos estudos analizados (Anderson e Freebody, 1983; White, Wayne e Graves, 1989 e Meara e Buxton, 1987) son semellantes tanto no relativo ós obxectivos coma ós resultados. Dos tres se poden extraer similares conclusións acerca dos tests de resposta tipo Si/Non como alternativa á medición do tamaño do vocabulario, sobre todo fronte ós de elección múltiple, xa que a comparación con medidas externas resulta máis satisfactoria.

Así e todo, hai unha serie de cuestións que deben ser tidas en conta:

a) Utilizan diferentes métodos para determina-la validez concorrente, co cal é difícil realizar comparacións entre os instrumentos. Anderson e Freebody (1983) e Meara e Buxton (1987) obtiveron correlacións semellantes entre os resultados con tests Si/Non e de elección múltiple (0,84 e 0,703 co grupo total, respectivamente, e 0,83 utilizando un grupo con maior homoxeneidade, no caso de Meara e Buxton). O procedemento que utilizaron para valida-lo test cunha medida externa foi moi similar no estudo de Anderson e Freebody (1983) e no de White, Wayne e Graves (1989), pero, se ben os resultados indican que o test Si/Non funciona mellor, discrepan considerablemente nas correlacións entre a entrevista, a proba de elección múltiple e o test Si/Non, debido á diverxencia en canto ós criterios de selección das palabras usadas na entrevista. Así mesmo, Meara e Buxton obtiveron correlacións positivas

con medidas externas (*Cambridge First Certificate* e outras probas utilizadas para situar ó alumnado na rede *Eurocentres*), o que é outra evidencia da validez do test. Os resultados, en definitiva, parecen mostrar que os test Si/Non son máis estables nas diferentes mostras empregadas e producen unha sobreestimación do tamaño do vocabulario menor que as probas de elección múltiple. O procedemento utilizado nestes tests é máis eficaz que o utilizado nas probas de elección múltiple, porque permite corrixi-la tendencia a responder por conxectura: introduce a corrección da sinceridade.

b) A fórmula de corrección da sinceridade proposta por Anderson e Freebody non parece resultar moi acertada. Non discrimina adecuadamente entre suxeitos con puntuacións en falsas alarmas significativamente diferentes (Táboa 3.1.). Meara (1991, 1993a, 1994) tratou de utilizar un procedemento baseado nos principios da "Teoría de Detección de Sinais", aínda que con este procedemento en determinados casos obtéñense puntuacións negativas. Meara (1991) recoñece a dificultade para encontrar un sistema de corrección que permita solucionar estes problemas e suxire a utilización de medidas non paramétricas de discriminación, adoptando o índice de discriminación proposto por Pastorre e Scheirer (1974), cuestión que se explicará máis adiante na segunda parte desta tese (capítulo 4, punto 4.4). Non obstante, é preciso sinalar que máis recentemente o autor elaborou un procedemento de corrección que permite obter estimacións máis precisas (Meara, 2002).

Táboa 3.1.: Porcentaxe de palabras coñecidas a partir de Acertos e Erros (Anderson e Freebody, 1983)

A \ E	E			
	1	2	5	10
30	1,000	1,000	1,000	1,000
29	0,966	0,964	0,960	0,950
28	0,931	0,929	0,920	0,900
27	0,897	0,893	0,880	0,850
26	0,862	0,857	0,840	0,800
25	0,828	0,821	0,800	0,750
20	0,655	0,643	0,600	0,500
15	0,483	0,464	0,400	0,250

A= Palabras sinaladas correctamente. E= Pseudopalabras sinaladas.
 Entrar na columna da esquerda con A e buscar no cruce con E a proporción correspondente

c) En casos como no de Anderson e Freebody non está claro como levan a cabo a mostraxe das palabras para a estimación do vocabulario total, e noutros, como no de White, Wayne e Graves, a mostra non parece que sexa o suficientemente ampla.

En resumo, os tres estudos, presentan evidencia da validez dos test de tipo Si/Non, en tanto que permiten realizar estimacións axustadas do tamaño do vocabulario. Os tres estudos analizados mostran a superioridade destes tests sobre os de elección múltiple, á vez que se mostraron como suficientemente válidos utilizando diferentes mostras.

5. O *Eurocentres Vocabulary Size Test* (Meara, 1990; Meara e Jones, 1988, 1990).

Este test é unha versión informatizada dos tests de resposta Si/Non utilizada como unha medida do vocabulario total en inglés para situa-lo alumnado no nivel máis apropiado na rede *Eurocentres*. Para a súa elaboración, os autores seleccionaron unha mostra da lista das

10.000 palabras máis frecuentes en inglés. Dividiron as palabras en 10 niveis ou bandas de frecuencia, con 1000 por nivel ou banda. O primeiro nivel de estratificación contén as 1000 palabras máis frecuentes, o segundo as seguintes 1000 máis frecuentes, etc. A avaliación comeza co nivel de frecuencia máis alto. Para comprobar cantas palabras coñece un determinado suxeito, utilízase unha mostra de 10 palabras das 1000 de que consta ese nivel de frecuencia e 10 pseudopalabras. Se o suxeito puntúa alto nese nivel de frecuencia, pásase a comproba-lo seguinte nivel, e o proceso continúa ata que a puntuación é moi baixa, menor do límite establecido. Neste punto, efectúase unha estimación do número de palabras que coñece o suxeito. Por exemplo, se unha persoa acerta o 100% dos ítems dos niveis do 1 ó 4, pero só o 20% do quinto, o programa calcula que esa persoa coñece entre catro e cinco mil palabras. Seguidamente, lévase a cabo unha apreciación máis detallada do quinto nivel. Para elo, utilizan 1 palabra de cada 20 das 1000 de que consta o quinto nivel (50 palabras en total), que representan o 5% das 1000 palabras. Rematada esta avaliación, calcúlase a estimación final do tamaño do vocabulario.

As correlacións obtidas entre as puntuacións neste test e as puntuacións na proba *Joint Entrance Test (JET)*, de acceso á rede *de Eurocentres* de ensinanza de inglés a estranxeiros, son altas para diferentes grupos de estudantes con distintas linguas maternas (entre 0,556 e 0,790). Tendo en conta que a proba *JET* valora habilidades lingüísticas como comprensión e precisión gramatical, pódese considerar que tamén son altas as correlacións do test de vocabulario con estas habilidades lingüísticas.

Máis recentemente (Meara, 2005, por exemplo), elaborou unha nova versión (*X_LEX*) para diversos idiomas (inglés, español, portugués, francés e sueco). A proba está composta por palabras organizadas en cinco niveis de frecuencia. Permite obter un perfil da proporción de palabras que se coñecen en cada nivel, así como unha estimación do vocabulario coñecido en relación coa mostra de vocabulario incluída no test (de 0 a 5000), que é un vocabulario considerado básico. A xuízo de Meara, nesta versión persisten dificultades

relacionadas coa corrección das respostas, así como cuestións sen aclarar como, por exemplo, se o significado da relación entre un determinado nivel de rendemento no test e outras habilidades e competencias lingüísticas é igual ou diferente nos diversos idiomas en que se pode utilizar.

6. *Threshold Level Galego Vocabulary Tests* (García e Meara, 1996). Fundaméntase no test de Meara e Jones (1988). Esta proba ten como finalidade estima-lo tamaño do vocabulario de persoas que non son galegas ou cun limitado coñecemento desta lingua. Consta de catro tests en galego elaborados a partir do *Nivel Soleira* do Consello de Europa (Fernández, Monteagudo e Puga, 1993), que é un vocabulario básico para adultos que visiten Galicia temporalmente e pode ser considerado como un vocabulario nuclear en galego, xa que a maioría dos ítems son palabras de alta frecuencia pertencentes a un amplo rango de contextos. Os catro tests están compostos por palabras extraídas do *Nivel Soleira*, pero os autores non incluíron palabras das bandas de frecuencia máis altas. Por este motivo, é probable que suxeitos cun escaso coñecemento do galego non coñezan un número de palabras superior ó 40%, incluso se dominan o español, tendo en conta que hai palabras que ben son iguais, ben son moi semellantes en ámbalas dúas linguas.

Cada test consta de 40 palabras reais e 20 pseudopalabras. O suxeito pode completar dous, pero para incrementa-la súa confiabilidade, é preferible utiliza-los catro ou, cando menos, tres. As respostas contabilízanse como Acertos (palabras reais sinaladas) ou como Erros (pseudopalabras sinaladas) e a puntuación consúltase nunha táboa de dobre entrada de características semellantes á que figura no Anexo IV, na que se teñen en conta os Acertos e os Erros. A puntuación que proporciona a táboa representa a porcentaxe de palabras coñecidas no *Nivel Soleira*. Calcúlase a media das puntuacións obtidas nos distintos tests separadamente para determina-la puntuación total e, deste modo, é posible realizar unha estimación do tamaño do vocabulario, pero unicamente en relación co devandito *Nivel*

Soleira. No caso de que a dispersión das puntuacións dos diferentes subtests abrangan un amplo rango, é preferible rexeitar a que se desvíe máis e usa-las puntuacións dos outros tests.

É importante resaltar que as puntuacións non provén información sobre a proporción de palabras infrecuentes que coñece o suxeito en galego, porque, como xa se explicou, a fonte da que se seleccionaron as palabras para a construción do test está composta por palabras habituais na fala cotiá. Non obstante, a xuízo dos autores, posto que existe unha porcentaxe destas palabras infrecuentes que adoitan ser semellantes en galego e en español, os aprendices de galego poden identificar un certo número das mesmas, aínda que non sexa moi amplo o seu vocabulario en galego.

7. *EVOCA* (Suárez, Seisdedos e Meara, 1998). Preséntase aquí de forma sucinta este test, xa que se comentou con maior amplitude en qué consiste no punto 2.3.5 do capítulo 2. *EVOCA* é unha proba que primordialmente pretende efectuar unha estimación do vocabulario en castelán que posúe un escolar entre 8 e 16 anos, pero tamén permite establecer comparacións entre os escolares utilizando os baremos que proporcionan os autores.

Foi elaborado a partir dunha selección de 1954 palabras do *Diccionario Escolar Etimológico* de García Hoz (1989), que posteriormente foron categorizadas por tres xuíces en catro grupos segundo a súa dificultade. Tamén utilizaron unha mostra de 1000 pseudopalabras das que seleccionaron 200. Consta de dous tests, con cinco subprobas cada un. Á súa vez, cada subproba consta de 60 ítems (40 palabras e 20 pseudopalabras). A tarefa consiste en que o suxeito sinala aquelas palabras que coñece.

A fiabilidade foi calculada coas puntuacións de Acertos utilizando o procedemento de dúas metades (elementos pares e impares), obtendo uns coeficientes de 0,86 para a Batería 1 e de 0,87 para a Batería 2. Para calcula-la fiabilidade, utilizouse, ademais, o procedemento de formas paralelas, obtendo unhas correlacións entre os diferentes subtests que oscilan

entre 0,72 e 0,78 para as puntuacións de Acertos e entre 0,51 e 0,74 para as puntuacións de Acertos menos Erros.

No que se refire á validez, os autores consideran que a de construto apóiase no proceso de construción do test. A validez criterial e concorrente apóiana, por unha parte, na capacidade do instrumento para discriminar entre grupos de escolares de distintas idades e entre os niveis socioeconómicos que establecen, e, por outra parte, nas correlacións coas puntuacións noutra proba de vocabulario como é a TEA-V, que van desde 0,44 a 0,53.

Como conclusión en relación coas características dos tests de resposta Si/Non, cabe dicir que a aplicación da "Teoría de Detección de Sinais" parece ter favorecido o desenvolvemento de tests que semellan alcanzar unha considerable utilidade. Ora ben, presentan algunhas limitacións como a xa mencionada sobre que non sempre resultan idóneos para avalia-lo coñecemento do vocabulario despois dun período de instrución, debido a que os suxeitos poderían responder máis por familiaridade que polo coñecemento do significado.

Outra limitación fai referencia a que aparentemente só son de utilidade para avalia-lo coñecemento pasivo, xa que se se compara a cantidade de vocabulario estimado con tests Si/Non e o estimado coa escala de Dale (1965), que se describirá no punto 3.3.1, os primeiros só valoran o coñecemento ata a segunda etapa da devandita escala, que reflicte unicamente o recoñecemento da forma, coñecemento receptivo, pero pode non implicar un coñecemento do significado. Ó respecto, Meara (1990) aduce, primeiro, que non hai nada equivocado na medida do vocabulario pasivo por recoñecemento, xa que é a habilidade básica na que se sustentan tódalas demais; segundo, que hai unha relación obvia entre tamaño de vocabulario pasivo e activo que, por outro lado, varía sistematicamente; e terceiro, que o coñecemento que avalían os tests Si/Non é menos pasivo do que parece, xa

que coa utilización de pseudopalabras, que son posibles palabras reais, próbase o grao de seguridade sobre a habilidade para usar palabras que o suxeito di coñecer.

En canto á cuestión da relación entre tamaño e profundidade do coñecemento do vocabulario, por exemplo, Quian (1998) estudou esta cuestión con adultos coreanos e chinos ($N=74$) que aprendían inglés e obtivo unha correlación entre ámbolos dous aspectos de 0,82. O tamaño vocabulario foi valorado cunha proba na que a tarefa basicamente consistía en emparellar seis palabras estímulo con tres definicións. A profundidade valorouna cunha proba que esixía relacionar por sinonimia unha palabra estímulo coas que figuraban nun grupo de catro e coas doutro grupo de catro pola súa posible relación de coocorrenza (colocación). Obtivo, ademais, unha correlación de 0,63 utilizando outra proba para valoralas profundidade complementaria da anterior, consistente en definir palabras estímulo ou traduci-las á lingua materna propia e en discernir se a palabra cambia de categoría gramatical ó engadir-lle un afixo. Concluíndo, a evidencia proporcionada por esta investigación parece indicar que o tamaño do vocabulario é un indicador fiable da súa profundidade, cando menos como o mide o test utilizado nesta ocasión.

Unha terceira limitación está relacionada coa utilización de pseudopalabras. Por un lado, segundo Meara (1994), podería mellorarse o funcionamento dos tests de tipo Si/Non se se utilizase un maior número de pseudopalabras, pero a tarefa sería excesivamente ardua. Por este motivo, decidiuse incluír a metade de pseudopalabras que de palabras. Outra cuestión é en qué medida afecta ás puntuacións a natureza deste tipo de ítems (pseudopalabras). Meara (1990) achega dúas respostas ó respecto. Unha é que os tests funcionan de forma distinta con grupos lingüísticos distintos, de modo que as pseudopalabras que son plausibles para uns grupos parecen non selo para outros. O segundo argumento de Meara é que os suxeitos adoitan ser prudentes no número de pseudopalabras que sinalan; alén disto, a tendencia a sinalar pseudopalabras pode ser reducida coas instrucións de administración da proba.

En consecuencia, por todo o sinalado, estes tests son susceptibles de melloras, é preciso levar a cabo investigacións sobre a súa fiabilidade e validez e, polo tanto, non deben ser entendidos como unha solución definitiva, aínda que como se viu teñen xa unha considerable tradición.

3.3. Medidas cualitativas ou de profundidade

Os instrumentos analizados ata agora teñen como finalidade estima-lo tamaño do vocabulario e son útiles en canto á medición da proporción de incremento do vocabulario que se produce, ben espontaneamente, ben como resultado da aplicación de programas de instrución, aínda que máis escaso neste segundo caso, como tamén son útiles para a avaliación das diferenzas individuais.

Pero estas medidas non permiten abordar cuestións máis finas relacionadas co significado das palabras, o seu uso apropiado, o coñecemento de que palabras poden ir asociadas ou coocorrer con outras (colocacións), os sinónimos, a distinción entre diferentes significados dunha palabra, etc., agás en tests coma a proba de "Vocabulario" do WISC, que dalgún modo permiten unha avaliación cualitativa. Por outro lado, como se analizou no capítulo 1, non é unha cuestión sinxela determinar que significa coñecer unha palabra, xa que o coñecemento de palabras desenvólvese ó longo do tempo, o seu significado pode enmarcarse nunha complexa rede de asociacións, significados, usos, etc., e non é posible dicir que se coñece ou descoñece unha palabra en termos absolutos, senón que tamén poden ser coñecidas de forma incompleta. Por elo, parece necesario desenvolver instrumentos de medida que capten os diferentes niveis do coñecemento dunha palabra.

Debido ás razóns que se acaban de comentar, o interese pola avaliación do vocabulario non se limita á medición do tamaño, senón que tamén se centra na avaliación da calidade do mesmo. Existen diversas escalas, que reflicten, cando menos, algunhas das características que definen de forma máis ampla o coñecemento dunha palabra, que

propoñen modos de avaliación diferentes dos compoñentes do coñecemento da palabra e, nalgúns casos, interpretalos por medio da identificación de diferentes etapas. Pero, a tarefa que esixen estes instrumentos de avaliación é ardua, xa que requiren moito tempo para a aplicación, son pouco económicos e limitan o alcance do test a poucas palabras. Por outra parte, estas escalas son difíciles de interpretar, principalmente porque falta un coñecemento sobre o desenvolvemento das etapas que dificulta a construción de escalas lineares.

A continuación, utilizando como base a información que achega Vives (1995), comentaranse algúns intentos de identifica-los compoñentes do coñecemento da palabra e algunhas escalas destinadas a capta-las diferentes etapas deste coñecemento. Como se sinalou anteriormente, os principais problemas que xorden ó respecto son a multidimensionalidade do mencionado coñecemento da palabra e a dificultade para establecer etapas progresivas ou lineares que reflectan unha orde de adquisición. Inclúese este contido coa intención de complementa-la discusión sobre a dificultade para decidir que significa coñecer unha palabra, cuestión xa analizada no capítulo 1, aínda que non atinxa ó tratado nesta tese de forma estrita. É a posibilidade dun coñecemento parcial do significado das palabras o que serve para entronca-lo contido deste apartado co obxecto da tese: construír un test que permita estima-lo tamaño do vocabulario coñecido, sen prexulgar que tipo e grao de coñecemento ten o suxeito das palabras sometidas a comprobación.

3.3.1 Escala de Dale (1965)

Este autor propón utilizar cinco graos na avaliación do coñecemento de palabras con falantes nativos dunha lingua, graos que se reflecten en cinco etapas, a última das cales trata de capta-la distinción entre palabras estreitamente relacionadas, pero que teñen significados diferentes (Figura 3.5.).

As etapas mencionadas reflecten aspectos do coñecemento das palabras que abranguen o descoñecemento (Etapa 1), o recoñecemento da forma (receptivo), que se

corresponde coa Etapa 2, o recoñecemento do significado en contexto (receptivo), que se corresponde coa Etapa 3, o coñecemento dun significado cando menos (Etapa 4) e o coñecemento de máis dun significado (Etapa 5).

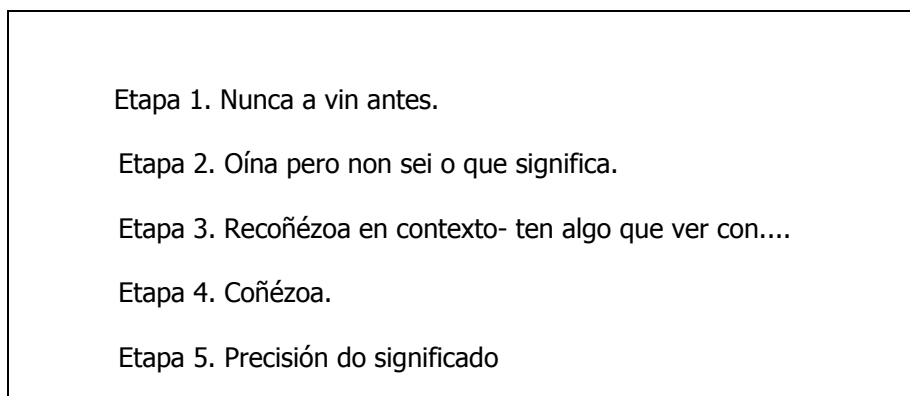


Figura 3.5.: Etapas no coñecemento dunha palabra (Dale, 1965)

A escala tamén pode ser aplicada para valora-lo coñecemento dunha segunda lingua, pero, con certas limitacións, xa que a orde das etapas non necesariamente sería a mesma. Representa un primeiro intento para valora-lo coñecemento das palabras nunha escala de desenvolvemento, aínda que se centra na forma e o significado, sen abordar outras características dese coñecemento.

3.3.2 Proposta de Read (1987)

Read centrouse na profundidade do coñecemento das palabras, entendida coma unha valoración dos diferentes compoñentes do devandito coñecemento, atendendo á súa multidimensionalidade. Desenvolveu un procedemento de entrevista para avalia-la aprendizaxe dunha segunda lingua, que se basea nunha lista de 30 nomes e 30 verbos extraídos da *University Word List* (Xue e Nation, 1984), e que presenta en forma dunha lista de control. Os suxeitos avaliados teñen que sinalar para cada palabra en que grao a coñecen nunha escala de tres alternativas (*Si/Non seguro/Non*). Obtense información adicional de

cada palabra mediante a resposta a diversas cuestións que inciden sobre a habilidade para pronuncia-la palabra, explicar que significa, identifica-lo campo de estudo, coñecer colocacións e asociacións e sinónimos.

Na análise dos datos, encontrou catro tipos de respostas, pero, en termos globais, é a primeira delas a que se utiliza de forma preferente:

- Definicións analíticas.
- Exemplos específicos de como é usada a palabra.
- Colocacións comúns ou palabras que coocorren habitualmente.
- Identificación da especialidade académica a que pertence.

Este foi un estudo a nivel exploratorio, que intentaba elaborar instrumentos que serviran para mostra-lo coñecemento das palabras nun abano de compoñentes como o concepto que representan, o seu uso, palabras coas que poden coocorrer e o campo semántico que abrangue. A preferencia por definicións analíticas mostra certa tendencia a centrarse no concepto, o que pode deberse á idade dos suxeitos avaliados e tamén ó tipo de instrución recibida. Os resultados están en consonancia coas descrições clásicas do coñecemento de palabras (Nation, 1990; Richards, 1976).

3.3.3 Vocabulary Knowledge Scale (Wesche e Paribakht, 1993, 1996)

Esta escala representa un novo intento de medi-la profundidade do coñecemento do vocabulario, entendendo este coñecemento coma un continuum, tendo en conta a súa multidimensionalidade, que abarca aspectos como a forma, diferentes significados, palabras asociadas ou que coocorren, derivación, regulamentos gramaticais, etc., o que, como xa foi comentado anteriormente, dificulta a utilización dunha escala linear. Combina ítems de autoinforme, nos que o suxeito debe sinalar en que medida coñece a palabra, con outros nos que o suxeito debe mostra-lo coñecemento da palabra. Deste modo, avalíase tanto a autopercepción do suxeito respecto do seu coñecemento coma o coñecemento que

realmente demostra con cada palabra estímulo, conforme unha escala de cinco niveis (Figura 3.6). A escala representa os cinco niveis de coñecemento que aparecen na Figura 1.4. (p. 30).

- 1: Non lembro ter visto a palabra antes.
 - 2: Teño visto a palabra pero non recordo o que significa.
 - 3: Teño visto a palabra e penso que significa ... (sinónimo ou tradución).
 - 4: Coñezo a palabra. Significa ... (sinónimo ou tradución).
 - 5: Podo usa-la palabra nunha frase.

Figura 3.6.: Escala de valoración do coñecemento dunha palabra (Wesche e Paribakht, 1996)

As respostas son puntuadas de acordo coa categoría máis alta conseguida. Respecto dos dous primeiros niveis, non hai modo de comproba-la veracidade da resposta, xa que só discriminan entre o recoñecemento ou non da palabra como unha forma do propio idioma. Nas demais categorías é necesario demostrar que realmente se coñece a palabra mediante a resposta ó ítem.

Nun estudo con 17 estudantes de inglés como segunda lingua, as correlacións obtidas por Wesche e Paribakht entre as categorías sinaladas do autoinforme e a utilización de palabras extraídas de catro temas traballados en clase, foi sempre superior a 0,95. Noutro estudo test/retest, con 93 estudantes de inglés como segunda lingua, dunha lista de 32 palabras non relacionadas con temas traballados, unha vez eliminada a primeira categoría (algúns aínda respondían nesta), a correlación para palabras contido foi 0,89 e para as palabras función de 0,82, polo que se pode considerar que se conseguen respostas fiables co instrumento.

A validez de construto da proba infírese do feito de que é sensible ás ganancias despois de instrución. A validez concorrente foi determinada a partir das correlacións coas puntuacións obtidas no *Eurocentres Vocabulary Size Test* de Meara e Jones (1990), a cal, como esperaban os autores, foi moderada (0,53), xa que miden aspectos diferentes e baséanse nun corpus diferente.

A partir do estudo, os autores conclúen que a escala permite discerni-lo grao de coñecemento representado polas etapas e é suficientemente precisa para detecta-las ganancias tras un período de instrución. Amais, ofrece a posibilidade de verifica-lo coñecemento real fronte ó percibido, a diferenza dos tests de tamaño do vocabulario.

Segue fielmente as etapas propostas por Dale e convérteas nunha escala de puntuación para cuantifica-lo coñecemento; non obstante, hai unha variación na primeira etapa respecto da Escala de Dale ("Non coñezo a palabra en absoluto" fronte a "Non recordo ter visto a palabra antes"), con elo, os autores intentan dar conta da diferenza entre a primeira vez que se encontra unha palabra e a primeira vez que se da conta da súa existencia, despois de estar exposto a ela varias veces, aínda que con elo non achega nova información, xa que nos dous casos a palabra non é familiar.

Por outro lado, o procedemento presenta unha serie de limitacións. Como xa se indicou, as dúas primeiras categorías só diferencian entre recoñecemento e non recoñecemento, polo que non é posible controla-lo fraude na resposta. O máis probable é que se sinale a segunda categoría, co que a primeira non é moi útil. A crítica que lle fan ós tests tipo Si/Non estes autores, mantense na súa proposta, aínda que os de resposta Si/Non controlan a sinceridade. Tampouco parece moi relevante pescudar se non recorda a palabra ou se a recorda, pero non sabe o significado. As categorías tres e catro introducen elementos que permiten asegurar que se coñece a palabra, que discriminan entre a certeza e a prudente probabilidade, pero cabe a posibilidade de responder á cuarta e se se equivocan non son penalizados. Realmente non achega máis información que outra que

esixa, por exemplo, escribir un sinónimo para cada palabra ou traducir e escribir unha frase e que se podería puntuar nunha escala de 0 a 3 coma a que aparece na Figura 3.7.

0. Non resposta ou resposta equivocada.

1. Sinónimo ou tradución correcta.

3. (Se é correcto) A palabra é apropiadamente usada nunha frase.

Figura 3.7.: Niveis “reais” de coñecemento dunha palabra na Escala de Wesche e Paribakht (1996)

A tarefa resulta moi ardua tanto para a persoa avaliada coma para a que examina, co que non compensa en relación coa información que proporciona, xa que, ó ter que ser contrastadas as respostas, limita o número de ítems e de suxeitos por test para que sexa manexable. As categorías de autoinforme parecen superfluas ó non poder corrixi-la sinceridade ou ter que facelo con exemplos ilustrativos de cada ítem. Tamén está limitada no sentido de que só permite avaliar un significado por palabra.

En canto á fiabilidade test/retest, aínda que as correlacións son altas, debido a que a mostra é reducida, sería interesante realizar réplicas con tests que contivesen outras palabras e comparar con tests previamente estandarizados, xa que a comparación co *Eurocentre Vocabulary Size Test* non parece que achegue datos concluíntes. A validez de construto baséase en que o test foi sensible ás ganancias en vocabulario despois dun período de instrución nun estudo piloto con 17 suxeitos e noutro con 38 suxeitos. Pero os autores non achegan datos concretos nos traballos consultados (1993, 1994, 1996). Tampouco utilizan outro tipo de procedementos, como a comparación con outros tests de vocabulario cos que se avaliasen os mesmos temas.

Finalmente, a conclusión dos autores de que, por unha parte, capta a progresión no coñecemento só se basea nun estudo piloto e os datos non parecen concluíntes. Por outro lado, a combinación do autoinforme e demostración do coñecemento, se o autoinforme se considera fiable, resultaría redundante e incrementaría a carga de traballo. A evidencia que ofrecen tampouco permite deduci-la viabilidade de usa-las categorías de autoinforme como puntuacións en futuras versións do test.

3.3.4 Associates Vocabulary Test (Read, 1993)

Supón unha nova versión do *Associates Vocabulary Test* (1990), co obxectivo de desenvolver unha medida do coñecemento do vocabulario que dea indicacións sobre a profundidade do coñecemento das palabras. A partir de catro tipos de tests existentes (de tamaño, verificación de datos, presentación de palabras en contexto versus illadas e de profundidade no coñecemento) o autor intentou busca-la fórmula que permitise utilizar equilibradamente estes elementos.

No deseño de 1993, a descrición do coñecemento dunha palabra adoptouna da evidencia da investigación sobre asociacións de palabras. En lingua materna adoitan producirse seguindo patróns estables que son interpretados como un reflexo da organización léxica. No caso dunha segunda lingua difire o tipo de asociacións, máis baseadas nas semellanzas fonolóxicas (asociacións *clang*), aínda que a medida que se incrementa o léxico tamén as asociacións van aproximándose ás producidas polos falantes dunha lingua materna, cando menos a nivel receptivo.

Cada ítem do test consiste nunha palabra estímulo e outras oito palabras, catro das cales son asociables. A tarefa consiste en identificalas, asumíndose que a selección de asociados está relacionada coa profundidade do coñecemento do vocabulario. As palabras asociables representan diferentes matices no significado da palabra estímulo e ilustran diferentes tipos de vínculos semánticos (paradigmático, sintagmático e analítico). A proba

elaborouse extraendo as palabras dun corpus de 800 da *University Word List (UWL)* compilado por Nation (1990, pp. 235-239), coa finalidade de medi-la aprendizaxe de estudantes universitarios en inglés (Figura 3.8.).

Ítem 50			
denominator			
common	develop	divide	eloquent
fraction	mathematics	species	western

Figura 3.8.: Ítem número 50 de *Associates Vocabulary Test* (Read, 1993).

O procedemento de puntuación non está claramente descrito, pero a cada ítem asígnanselle un máximo de catro puntos se o suxeito sinala as catro opcións asociadas correctamente, sen contempla-los erros nin penalizalos, polo que non hai modo de controla-la sinceridade e, ademais, foméntase que os suxeitos emitan algunha resposta ante a dúbida.

A fiabilidade investigouse correlacionando as puntuacións de dúas formas paralelas do test (0,97). Para a validez concorrente, correlacionou os resultados nas dúas formas do seu test cos acadados na batería *English Proficiency Test*, incluído o test de vocabulario, obtendo unhas correlacións de 0,76 e 0,81.

A pesar de que a fiabilidade é moderadamente alta, parece que se confunden a medida da profundidade do coñecemento do vocabulario coa habilidade dos suxeitos para adiviñar ou supoñer-la resposta, xa que a identificación de palabras asociadas non necesariamente mostra que a palabra estímulo sexa coñecida. Pode que a resposta resulte da combinación de varios factores, como bo coñecemento xeral do vocabulario, recursos

para buscar posibles asociados e confianza nas propias suposicións, como suxire o propio autor.

Existen, por outra parte, cuestións por resolver que obrigan a adoptar precaucións en relación coa fiabilidade e a validez concorrente: o método de puntuación, o papel do contexto e o número de palabras que compoñen cada ítem. A primeira refírese á falta dun método adecuado de puntuación que permita verifica-las respostas, posto que o que se usa está moi afectado pola utilización de estratexias baseadas na adiviñación e repercute negativamente na fiabilidade e na validez. Só se contabilizan as palabras asociadas sinaladas, mais non se controlan o número de *distractores* ou alternativas incorrectas sinaladas nin o número total de palabras seleccionadas na resposta. Así, se unha persoa sinala tódalas palabras, obtense unha puntuación alta de forma nesgada e non é posible diferenciala da doutra persoa que marque só as catro correctas, é dicir, prodúcese unha sobreestimación. Outra consecuencia da estratexia de suposición é que provoca dúbidas sobre a validez de construto, porque cabe a posibilidade de responder en base ós vínculos semánticos entre os asociados obviando a palabra estímulo cando é descoñecida, o que leva a que se poda responder correctamente, sen coñece-la palabra estímulo, co que o obxectivo do test desvirtúase.

Outro problema está relacionado co contexto, xa que ó ser moi rico neste caso (hai moitas palabras), en tanto que é unha tarefa receptiva, redúcese a posibilidade de avalia-lo coñecemento dunha determinada palabra; o contexto ofrece excesivas claves e leva a sobreestima-lo vocabulario.

Unha terceira cuestión ten que ver co número de elementos lexicais en cada ítem. Este número parece demasiado elevado o que comporta efectos non desexables como é unha sobrecarga da tarefa, en tanto que esta tarefa de recoñecemento vese dificultada pola produción de múltiples hipóteses na mente ó ter que considerar tantas palabras. Tamén converte a elaboración do test nunha tarefa complexa, xa que as palabras deben ser de

semellante frecuencia á palabra estímulo e é necesario equilibra-los tipos de palabras. O elevado número de palabras contribúe, asemade, á heteroxeneidade da estrutura dos ítems e á variabilidade nos patróns de resposta.

3.3.5 Three Words Association Vocabulary Test (Vives Boix, 1995)

Este test de coñecemento do vocabulario en profundidade xorde para mellora-los intentos anteriores (os de Read en particular), pero ten en conta igualmente os tests de tamaño do vocabulario do tipo Si/Non de Meara.

Para a súa elaboración, a autora levou a cabo sete estudos en lingua catalana e castelán con diferentes mostras, compostas por suxeitos desde bacharelato á universidade e adultos, que tiñan como lingua materna o catalán, o español, o inglés e o francés. Os tests foron construídos seleccionando palabras de *POAL. Vocabulari bàsic ampliat* (1978) e de *Nivel Llindar* (1986) para o catalán e do dicionario de Rodríguez Bou (1952) para o español. Os tests variaron en canto ó número de ítems, pero o definitivo consiste basicamente nunha lista de pares de palabras que poden ter un vínculo asociativo entre elas ou non. O suxeito avaliado debe recoñecer se hai ou non relación entre as dúas de cada par. Se a resposta é afirmativa, debe marca-lo par no espazo destinado a este fin. Se é negativa ou non está seguro, debe deixa-lo espazo en branco.

Adapta o formato ó dos test de resposta Si/Non. En lugar de ítems dunha soa palabra, presenta pares de palabras, asociadas ou non, e non hai pseudopalabras. En lugar de *distractores* hai pares sen relación obvia. Nas Figuras 3.9. e 3.10. ofrécese mostras de dúas versións do test utilizado na investigación (a primeira en catalán e a segunda en castelán, nas que as instrucións foron traducidas por nós ó galego). A tarefa resulta máis esixente que a dos tests Si/Non e permite obter gran cantidade de información sobre o coñecemento da palabra en cada ítem. A asunción básica é que canto mellor se coñeza unha palabra, máis probable é que recoñeza unha das asociadas.

GV92A

O que tes que facer:

- Le-la lista de parellas de palabras atentamente.
- Se pensas que están relacionadas semanticamente, escribe a resposta [Si] no espazo en branco.
- Se pensas que non están relacionadas, ou non o sabes, pon unha cruz [X].

1 [] mudarse : pis

2 [] escampar : boira

3 [] oli : sal

4 [] actor : teatre

59 [] frase : fre

60 [] tarda : verí

Figura 3.9.: Test *AVT* utilizado no estudo piloto en catalán

GV94G

En cada ítem hai dúas palabras que están fortemente relacionadas entre si, mentres que unha terceira non o está. Sinala cunha cruz a que non está relacionada coas outras dúas.

Non deixes ningún espazo baleiro.

1 acceptable

válido

pitar

41 lisonja

codo

alabanza

120 leyenda

pronosticar

tiempo

Figura 3.10.: Test *A3VT* definitivo en castelán

Cabe sinalar que, a xuízo da autora, este test presenta varias vantaxes respecto a intentos anteriores:

a) Pola simplicidade do seu formato, é fácil de administrar coma o son os tests de resposta Si/Non, e, así mesmo, pódese aplicar e corrixir de forma informatizada.

b) É unha medida do tamaño do vocabulario, en tanto que permite probar un amplo número de palabras nun corto período de tempo. Propóñense 120 palabras en cada test nunha batería de tres tests, co que se utilizan 360 palabras.

c) É unha medida do coñecemento do vocabulario, posto que avalía este coñecemento de forma máis fonda que os tests Si/Non, xa que non só demanda recoñecemento das palabras senón os significados das mesmas e a relación entre elas. Neste sentido o coñecemento non só é pasivo, xa que é preciso activa-las palabras de cada ítem, de modo que se pode considerar case produtivo.

Mentres que a xuízo da autora os tests Si/Non só discriminan entre as dúas primeiras etapas da escala de Dale (punto 3.3.1 deste capítulo), relativa ó recoñecemento dos atributos físicos da palabra, o seu pode discriminar entre as etapas cuarta e quinta. Non só é necesario coñece-lo significado das palabras, senón tamén o significado particular que fai que a asociación sexa posible; necesítase bastante certeza sobre os diferentes significados e das combinacións posibles; permite valora-lo coñecemento de diferentes significados dunha palabra e, no caso das polisémicas, pódense presentar en tantos contextos de asociacións como demanden os diferentes significados e usos, polo que permite adaptarse a diferentes niveis de competencia.

d) Utiliza un método útil de puntuación, xa que, coma nos tests de resposta Si/Non, aplícase a “Teoría de Detección de Sinais” que posibilita a corrección da sinceridade, co fin de calcula-lo coñecemento real.

e) Presenta a vantaxe de non utilizar pseudopalabras e, deste modo, permite que o número de palabras avaliadas sexa maior, á vez que evita que se produza unha aprendizaxe incidental das pseudopalabras.

Nos dous primeiros experimentos a autora intentou explora-la fiabilidade e os resultados mostran índices positivos, pero moderados. Prodúcese unha variabilidade entre tests e os resultados máis fiables obtéñense cos suxeitos de maior nivel lingüístico. Este patrón de resultados dase aínda cando se controlan factores como os cognados, a frecuencia de palabra e a orde do test.

No terceiro estudo, Vives investigou o efecto do nivel lingüístico, do método de puntuación e do tamaño do test na variabilidade entre diferentes formas do test e o modo de incrementa-la fiabilidade. Comprobou que controlando as variables mencionadas e utilizando a fórmula estándar de corrección, conséguense niveis aceptables de fiabilidade, xa que a “Teoría de Detección de Sinais” penaliza excesivamente a sinceridade, sobre todo no caso dos suxeitos que puntúan baixo.

O cuarto estudo estivo destinado a investiga-la validez concorrente, que a autora estableceu comparando coa habilidade de tradución medida cunha proba elaborada ó efecto. As correlacións son significativas, pero moderadas, o que é atribuído ás diferenzas no nivel de esixencia entre o *Asociates Vocabulary Test* e a proba de tradución, que a autora considera como máis sinxela. Unha segunda razón sería o número de ítems, confirmándose que non debe ser inferior a 100 para poder obter correlacións altamente significativas e poder efectuar estimacións fiables. Outra explicación é que se debe a que se produce unha perda de información que provoca unha subestimación do tamaño do vocabulario, sobre todo no caso de suxeitos con baixas puntuacións. Isto débese a que se poden dar situacións nas que a persoa avaliada coñeza unha das palabras ou as dúas, pero non a relación entre elas. Para intentar salvar estes obstáculos, a autora elaborou unha forma alternativa do test, o *A3VT* con 120 ítems, divididos en tres bloques de 40 ítems cada un. Levou a cabo tres

experimentos e os resultados amosan que é útil como medida da organización léxica e que correlaciona alto con outras medidas de vocabulario como é a tradución.

Aínda que o obxectivo era elaborar unha medida tanto do tamaño coma da profundidade do vocabulario, esto mostrouse como algo moi complicado cun só test e o test *A3VT* non parece que permita efectuar estimacións precisas do tamaño. Si informa sobre a organización lexical (avalía por recoñecemento o que se podería considerar como a unidade mínima da organización léxica, a relación entre dúas palabras).

A fiabilidade é moi satisfactoria, obtida tanto con análises clásicos coma utilizando o modelo "Item Response Theory" de Rasch. A alta correlación coa tradución interpretase como evidencia da validez concorrente.

Con este enfoque, a descrición do lexicón non se basea só no coñecemento de palabras illadas, senón que intenta describi-lo sistema no nivel relacional, coma unha rede que se organiza de forma diferencial en falantes nativos dunha lingua e aprendices dela como segunda lingua. Pode significar unha boa contribución en aspectos como ó entendemento do desenvolvemento do lexicon mental, á súa avaliación comparando falantes nativos dunha lingua con outros que teñen esa lingua como segunda ou á avaliación do efecto de diferentes métodos de ensino.

A linguaxe verbal soamente a posuímos os humanos. A pesar das súas limitacións (a súa linearidade, por exemplo), permítenos construír un mundo mental moi sofisticado a partir dos estímulos que nos chegan a través dos sentidos. Este mundo mental, o pensamento, permítenos uns logros enormes en comparación con outros seres vivos (os logros da ciencia e da técnica, por exemplo).

No centro da linguaxe verbal aparece a palabra, que é máis fácil de definir como forma que como significado, como vimos nos capítulos que preceden. Realmente, é paradoxal con que presteza e seguridade falamos de forma cotiá da palabra (aí están os

dicionarios ...) e o difícil que resulta acoutala con rigor, cientificamente: dicir exactamente que é, decidir cando se coñece, contar cantos vocábulos coñece unha persoa, etc.

Con todo, pode considerarse suficiente o dito como contexto para pasar a tratar unha cuestión parcial, pero importante e oportuna, que constitúe a contribución empírica e máis persoal deste traballo, a segunda parte do mesmo: a construción dun instrumento que permita estima-lo vocabulario en galego que coñecen os escolares da nosa comunidade autónoma.

SEGUNDA PARTE

Construción dun test de tamaño do vocabulario en galego

Capítulo 4

Primeiras fases no proceso de construción do test

Introdución

Segundo Nation (1993a), unha das primeiras investigacións publicadas sobre o tamaño do vocabulario é o estudo de Kirkpatrick de 1891, no que valorou o seu propio coñecemento a partir do dicionario *Webster*. Desde entón, leváronse a cabo numerosos intentos de estima-lo volume de vocabulario na lingua inicial dos suxeitos utilizando dicionarios como punto de partida. Non obstante, Nation sinala que non se dispón de estimacións fiables da cantidade de vocabulario coñecido polos falantes dunha lingua, aínda que nas dúas últimas décadas se produciu unha reactivación do interese polo desenvolvemento de medidas estandarizadas.

Sería, polo tanto, de utilidade un método fiable no estudo de cuestións relacionadas co coñecemento e desenvolvemento do tamaño do vocabulario de falantes na súa lingua inicial ou de aprendices de L2, para determina-la ratio de aprendizaxe na adquisición de novas palabras ou para comparar diferentes tipos de programas de instrución (Meara, 1993b).

A partir da revisión da literatura sobre o tema, parece claro que a medida do vocabulario desenvolveuse seguindo dous enfoques, en función de que se lle concedese maior importancia á calidade e profundidade do seu coñecemento ou ó tamaño do mesmo. Esta división responde, non obstante, máis a razóns prácticas para manexa-la vasta dimensión e complexidade do coñecemento do vocabulario, que á necesidade en si de dividir este coñecemento en dous compoñentes. De feito, poderíase asumir que unha boa actuación nun test que explora o tamaño do vocabulario pasivo, pode ser indicativa do vocabulario

activo, é, á inversa, dunha boa actuación en canto á calidade cabe a posibilidade de inferir que a persoa dispón dun amplo vocabulario.

Na investigación sobre este tema detéctase, por outra parte, unha limitación das medidas xerais e comprensivas para capta-la complexidade do coñecemento do vocabulario. O primeiro grupo de traballos, os de tamaño, ofrecen tests con métodos de puntuación apropiados para contempla-la sinceridade, pero presenta limitacións que proveñen de que, en principio, a tarefa na que se basean é só de recoñecemento, non recollendo información sobre outros aspectos do coñecemento da palabra. No segundo grupo de traballos, baseados na calidade, non resulta sinxelo construír unha escala que capte a multidimensionalidade do coñecemento da palabra, que utilice un método de puntuación operativo, debido a que, coa utilización dun número reducido de ítems, dificilmente se pode levar a cabo unha estimación xeral da competencia en vocabulario.

Paul Meara desenvolveu unha liña de investigación baseándose na aplicación da “Teoría de Detección de Sinais” e sentou as bases para a estandarización dos denominados tests de resposta Si/Non. Conseguiu evidencia da viabilidade dun test de recoñecemento, que puntúa seguindo a devandita teoría, e que correlaciona significativamente con outras medidas da linguaxe.

Seguindo esta liña de investigación, a finalidade deste traballo, como xa se sinalou en diversas ocasións, consiste en construír, utilizando as achegas de traballos anteriores, fundamentalmente os de Meara (1993a, 1994), Meara e Buxton (1987) e de Suárez, Seisdedos e Meara (1998), un instrumento de resposta tipo Si/Non, o suficientemente fiable e válido, que permita, mediante cálculos baseados en inferencias, estima-lo tamaño do vocabulario galego coñecido polos escolares de Galicia entre 8 e 16 anos. Como tamén quedou sinalado anteriormente, estes tests esixen que a persoa avaliada responda "Si" no caso de que pense que coñece a palabra ou estímulo que se lle presenta, mentres que se

pensa que non a coñece ou non está segura debe responder con "Non" ou simplemente non emitir resposta algunha.

A explicación do proceso seguido para a súa construción, desde a extracción dunha mostra de palabras, pasando pola selección da mostra de suxeitos e o cálculo da fiabilidade e validez, ata a elaboración de baremos, constituirá o contido desta segunda parte da tese. Non obstante, antes de abordar esta tarefa, comentaranse algunhas cuestións en relación coa problemática que presenta a construción de tests de tamaño do vocabulario, para, a continuación, entrar na devandita explicación.

4.1 A problemática acerca da construción de tests de tipo Si/Non

Para a elaboración deste tipo de tests de vocabulario, o proceso adoita iniciarse coa selección dunha mostra de palabras, ben dun dicionario ou lista de frecuencias, ben, como é o caso nesta tese, dun dicionario xeral da lingua que se poida considerar que representa suficientemente o vocabulario das persoas que se van a examinar e da lingua acerca da que vamos a valora-lo vocabulario. Segundo Goulden, Nation e Read (1990) e vimos en 2.1.3, os estudos do tamaño do vocabulario en xeral presentan enormes diverxencias en canto ós resultados. Estas diverxencias prodúcense debido á falta de precisión á hora de resolver determinadas cuestións metodolóxicas, como son as que se comentan seguidamente.

4.1.1 Decidir que contar como unha palabra

No caso dos estudos baseados en mostras de dicionario, para efectua-la estimación, habitualmente lévase a cabo un cálculo do número de palabras do dicionario e posteriormente extráese unha mostra de palabras. Con este mostra avalíase o coñecemento que ten un determinado suxeito ou un grupo das palabras que a compoñen. A proporción de palabras que coñece interprétase como a porcentaxe de palabras que coñece de todo o

dicionario. Por exemplo, Diller (1978) asume que o *Webster's Third New International Dictionary* de 1961 contén 450.000 entradas, como se indica no propio dicionario. Seleccionou unha mostra de 1000 palabras e administroullas mediante un test a unha mostra de escolares, que obtiveron unha puntuación media de 480 palabras, o que representaba o 48% das do dicionario. Desta forma, calculou que o vocabulario total da mostra de escolares era de 216.000, o 48% das entradas do dicionario. Neste caso, debido a que non se contabilizou adecuadamente o número de palabras do dicionario, como se verá no que segue, a estimación do tamaño do vocabulario parece moi sobredimensionada.

Para que o cálculo sexa acertado, é esencial estima-lo número de palabras do dicionario do modo máis válido e fiable, xa que pode suceder como mostran, por exemplo, Goulden Nation e Read (1990), que, unha vez excluídos do dicionario *Webster* os nomes propios, palabras compostas e ítems como abreviaturas, símbolos, etc., redúcese o tamaño do seu vocabulario nun 50%. Isto implica decidir que conta como unha palabra e que non conta como unha palabra (por exemplo: abreviaturas, nomes de persoas, de lugares, outros nomes, prefixos, sufixos), así como cales se consideran *palabras base* e cales *derivadas*, xa que incluír na mostra as derivadas conduce a unha sobreestimación do vocabulario.

Nagy e Anderson (1984) e Goulden Nation e Read (1990) consideran que unha palabra base é a menos inflexionada do grupo de palabras relacionadas (morfema). Se a elección resulta moi complicada, pola semellanza en canto a lonxitude e inflexión, os nomes e verbos cóntanse como palabra en detrimento dos adxectivos ou adverbios (por exemplo, *congruencia* mellor que *congruente* ou *consecuencia* mellor que *consecuente*). Nation e Ellis (1993) propoñen que, cando sexa difícil decidir cal é a menos inflexionada, é preferible elixila primeira entrada, a non ser que na definición desta sexa usada a segunda como base (por exemplo, *reunir* mellor que *reunión* ou *líquido* mellor que *liquidez*), ou a primeira sexa moi pouco frecuente, casos nos que sería máis apropiado escolle-la segunda. No que se refire ós homógrafos, para Goulden Nation e Read (1990), é preferible escolle-lo primeiro. Se se trata

dunha palabra composta e o significado non é facilmente extraíble das que a compoñen, debería considerarse como palabra base (por exemplo, *cosmopolita* ou *cosmonauta*).

Estes mesmos autores, consideran como *palabras derivadas* aquelas nas que se pode extraer claramente o significado a partir dos seus compoñentes, as compostas por prefixos comúns e os homógrafos relacionados en canto a significado co que aparece como primeira entrada, mentres que son *palabras base* as que aparecen como entrada principal.

4.1.2 Elección das palabras para o test

É necesario seleccionar unha mostra representativa das palabras do dicionario elixido, porque, obviamente, sería imposible comproba-lo coñecemento de tódalas entradas que conteña o devandito dicionario. Pero, con obxecto de equipara-lo criterio entre diferentes investigacións, tamén é preciso excluír da elección as que non sexan palabras base, porque con frecuencia os dicionarios difiren considerablemente no modo en como tratan, por exemplo, as palabras compostas, no relativo á súa inclusión e lugar de inclusión.

Nation (1993a) propón o seguinte procedemento xeral, co fin de que a selección de palabras sexa o máis adecuada posible:

a) Elixir un dicionario coa suficiente amplitude para cubri-lo coñecemento do vocabulario das persoas a investigar. No ámbito anglosaxón, a partir dos cálculos elaborados por autores como Thorndike (1924), Lorge e Chall (1963), Goulden, Nation e Read (1990) ou D'Anna e Zechmeister (1991) e Nation e Ellis (1993), parece que en adultos cultos ou cun certo nivel de instrución o vocabulario base sitúase nunha media de arredor de 20.000 palabras, polo que o dicionario debería constar de cando menos 30.000 palabras.

b) Determina-lo número total de entradas do dicionario dun modo fiable, xa que é preciso coñecer este dato para efectuar unha estimación do tamaño do vocabulario e os autores, con frecuencia, son imprecisos ó respecto, como indican Goulden, Nation e Read (1990). Este cálculo pódese realizar manualmente (D'Anna e Zechmeister, 1991; Suárez,

Seisdodos e Meara, 1998), mediante procedementos informatizados (Nation e Ellis, 1993) ou mediante unha mostraxe do dicionario (Dupuy, 1974; Goulden, Nation e Read, 1990).

c) Expoñer claramente os criterios para decidir que ítems se excluirán e cales se considerarán como membros de familias de palabras. A principal razón para esta exclusión é a de evita-los posibles efectos que poidan ter sobre a estimación do vocabulario os criterios seguidos polos autores de dicionarios para a inclusión de palabras nos mesmos. Sirva como exemplo que no relativo ás entradas enciclopédicas (nomes, personaxes, lugares, plantas e fauna), non está claro cal sería o límite para a súa inclusión. Se se exclúen da mostra tipos de palabras que non tódolos autores inclúen nos dicionarios e se limita esta selección a aquelas palabras sobre as que hai acordo que poderían formar parte do vocabulario dunha persoa, a estimación do número de palabras que coñece un suxeito non se cinguiría exclusivamente ás referidas a un dicionario concreto, senón que podería ampliarse a outros dicionarios.

Outra razón de exclusión refírese a que canto máis baixa é a ratio *mostra de palabras extraída/número de palabras do dicionario*, máis sensible será a medida. Por exemplo, sería máis adecuada unha mostra elaborada extraendo 1 palabra cada 20 do dicionario que unha mostra elaborada a partir da extracción de 1 palabra cada 100 do dicionario. Non obstante, deste modo tamén será maior a mostra de palabras, o que incrementa a dificultade para levar a cabo unha comprobación do coñecemento das palabras da mostra nun tempo razoable. Por isto, é preferible reduci-lo tamaño do dicionario excluindo ítems (nomes propios, palabras compostas, abreviaturas, símbolos, partes de palabras, etc.).

Unha terceira razón, ten que ver coa finalidade de evita-la inclinación a elixir en maior medida palabras máis frecuentes, xa que existe unha relación entre frecuencia de ocorrencia e probabilidade de coñecela (Anderson e Freebody, 1983), á vez que tamén sucede que canto maior é a frecuencia dunha palabra, tamén o é o número de palabras relacionadas coa mesma. Os autores de dicionarios siguen diferentes criterios para incluí-las formas

relacionadas coa palabra base (listalos baixo a palabra base ou en entradas separadas), e tamén difiren en canto á inclusión de palabras que están compostas por outras xa incluídas con anterioridade.

d) Utilizar un procedemento de mostraxe que preveña unha elección desproporcionada dos ítems con maior número de entradas e que ocupan máis espazo:

Thorndike (1924) propón numerar as entradas do dicionario e extraer unha porcentaxe, como poden ser, por exemplo, 10 ó azar de cada 100 (así proceden Suárez, Seisdedos e Meara, 1998), eliminando homógrafos e derivadas. Existe igualmente a posibilidade de elixir un determinado número de entradas de cada páxina (procedemento usado por Goulden, Nation e Read, 1990), pero xera un problema de mostraxe, porque cada páxina contén diferente número de palabras e as que teñen poucas están sobrerrepresentadas na estimación do tamaño do vocabulario.

Outra posibilidade, consiste en efectuar un mostraxe ó azar, usando listas aleatorias (Dolby e Resnikoff, 1967) ou unha selección por métodos informatizados (Nation e Ellis, 1993). D'Anna e Zechmeister (1991) usaron mostras estratificadas baseándose nas letras do alfabeto, organizando a mostra de forma que represente o máis axustadamente posible a estrutura do dicionario, recollendo a mesma proporción de palabras que empecen con cada letra na mostra que no dicionario.

e) Elixir unha mostra o suficientemente ampla que permita estimar o tamaño do vocabulario con razoable nivel de fiabilidade.

O tamaño absoluto da mostra ten unha importante repercusión na fiabilidade da estimación do vocabulario, xa que ó realizala a partir dunha mostra de palabras comporta un erro; canto maior sexa a mostra, menor será este erro. Na Táboa 4.1. expónse a cuantificación teórica efectuada por Suárez, Seisdedos e Meara (1998, p. 21) do erro que se cometería na estimación do vocabulario utilizando distintas mostras de palabras, para tres niveis de confianza. Nation (1993a, p. 36) efectúa unha análise semellante.

Táboa 4.1.: Intervalos de confianza para diferentes tamaños de mostrás

Mostra de Palabras	Nivel de confianza		
	10%	5%	1%
100	8,2	9,8	12,9
150	6,7	8	10,5
200	5,8	6,9	9,1
300	4,7	5,7	7,4
400	4,1	4,9	6,5
500	3,7	4,4	5,8
600	3,4	4	5,3

Esto significa que, se, por exemplo, a un escolar se lle aplican 400 palabras extraídas dun dicionario de 20.000 e se estima que coñece 11.500 palabras, o número real estará determinado polo nivel de confianza co que se opere. Así, ó 10%, estaría comprendido entre 10.680 e 12.320 palabras; ó 5%, entre 10.520 e 12.480; e ó 1%, entre 10.200 e 12.800.

f) A mostra debería ser contrastada en canto á consistencia na aplicación do criterio de inclusión e exclusión dos diferentes tipos de ítems (palabras base, derivadas, etc.). Con este fin, cabe, por exemplo, elabora-la mostra por categorías e compara-la cantidade de palabras incluída en cada categoría, utilizar máis dunha persoa na súa elaboración e calcula-la correlación entre os ítems de diferentes clases seleccionados por cada unha delas ou comparar cunha lista de frecuencias.

g) A mostra debería ser contrastada cunha lista de frecuencias para asegurar que non se produza unha elección inadecuada dos ítems de maior frecuencia. Por exemplo, autores como Thorndike (1924), Williams (1932) e Lorge e Chall (1963) utilizan un procedemento segundo o cal se cada palabra da mostra representa 100 palabras do dicionario, haberá 10 palabras na mostra entre as 1000 palabras máis frecuentes na lista de

frecuencias (10 x 100), 10 nas seguintes 1000 máis frecuentes, etc. Goulden, Nation e Read (1990), utilizando un procedemento semellante, encontraron que era necesario volver a calcula-lo número de palabras nos diferentes niveis de frecuencia establecidos en Thorndike e Lorge (1944), co obxecto de evitar que nas mostras estean representados de xeito inadecuado os ítems de maior frecuencia, redefinindo o término *familia de palabras*, e incluído nel, amais da palabra base, os sufixos en *-s, -ado, -ando, -mente*.

h) É importante describir clara e explicitamente cómo se efectuou cada paso para permitir que o estudo se poida replicar.

4.1.3 Decidir como efectua-la avaliación do coñecemento do vocabulario

Entre as diferentes alternativas en canto ós test que existen para valora-lo tamaño do vocabulario coñecido por escolares, optamos, como xa quedou indicado reiteradamente, por un dos denominados de resposta tipo Si/Non, decisión que se adoptou tendo en conta tanto as vantaxes coma os inconvenientes que supoñen.

Tendo presente que non se trata de comproba-lo grao de coñecemento de palabras en particular, senón a habilidade de ser capaz de recoñecelas, é necesario, por unha parte, como sinalan Goulden, Nation e Read (1990):

a) Clarificarlle ós suxeitos se, para responder, deben preguntarse se viron a palabra antes, aínda que non coñezan o significado; se deben ser capaces de dar unha definición, ou que deben facer se dubidan.

b) Correxí-la probable tendencia a responder "Si". Para elo, autores como Anderson e Freebody (1983), Nagy, Herman e Anderson (1985), Meara e Buxton (1987), Meara e Jones (1988) ou Suárez, Seisdedos e Meara (1998) introducen pseudopalabras como ítems que permitan efectuar suposicións aceptables acerca de en que medida os suxeitos responden sen estar seguros de coñecer unha palabra. Neste sentido, asúmese que unha persoa que

responde “Si” a un número alto de pseudopalabras, probablemente será menos fiable nas súas respostas a palabras reais que outras máis escrupulosas.

Por outra parte, existen tres aspectos que é preciso ter en consideración, tanto na construción de tests coma na súa utilización:

a) Comprobase que poden ser significativamente diferentes as puntuacións de distintos subtests illados de 60 ítems, incluso cando a selección destes ítems se realiza totalmente ó azar. Unha das razóns desta variación é que o número de ítems que compoñen un subtest é excesivamente reducido en relación coa mostra total de palabras. Outra razón é que a puntuación de cada subtest é moi sensible ós erros, é dicir, ó número de pseudopalabras sinaladas, sobre todo se o número de acertos é baixo. Todo isto induce a pensar que a selección dos ítems debe ser coidadosa, xa que pode afectar ás puntuacións obtidas.

b) En certo número de casos, que se cifra entre o 5% e o 10%, a puntuación sitúase por debaixo de 1, debido a que o número de *Acertos* é menor ou escasamente maior que o de *Erros*, de forma que o sistema de corrección produce un resultado dificilmente explicable. Sucede como se estas persoas subestimaran o seu coñecemento das palabras e sobreestimaran o das pseudopalabras. Esta é unha cuestión sen resolver.

c) Un suposto no que se basan este tipo de test é que os *Erros* cometidos por un suxeito son un reflexo das suposicións que aplica nas palabras que acerta. Así, un alto número de *Erros* significaría que os *Acertos* son unha sobreestimación do coñecemento que realmente se posúe. Non obstante, cando se utiliza con suxeitos de determinadas L2, esta relación pode verse alterada, na medida en que poden influir-los erros de lectura ou as semellanzas ortográficas entre palabras dos dous idiomas, o que, no caso de tratarse dunha valoración de persoas en L1, pode implicar unha limitación cando os suxeitos presentan dificultades de lectura.

4.2 Elección dun dicionario

4.2.1 Características xerais dos dicionarios

A palabra dicionario presenta dúas acepcións segundo a Real Academia Española (2001, p. 818). Na primeira defíneo como "Libro no que se recollen e explican de forma ordenada voces dunha ou máis linguas, dunha ciencia ou dunha materia determinada". Na segunda como "Catálogo numeroso de noticias importantes dun mesmo xénero, ordenado alfabeticamente". A Real Academia Galega, pola súa parte, defíne o dicionario como "Un libro que contén todas ou parte das palabras dunha lingua ou dunha ciencia ou materia determinada, ordenadas alfabeticamente e acompañadas dunha definición ou doutro tipo de explicación" (1997, p. 413).

Realmente os dicionarios nunca recollen tódalas palabras ou voces dunha lingua porque é imposible na práctica. En palabras dun personaxe de Saramago (2003, p. 143) "Tódolos dicionarios xuntos non conteñen nin a metade dos términos que necesitaríamos para entendernos uns ós outros". Prodúcese, de feito, segundo Alvar (1993, p. 62) "Unha especialización das obras lexicográficas, non expresada (...) e que se debe sinalar á hora de diferenciar tipos de dicionarios". A xuízo deste autor, o dicionario da RAE é un dicionario da lingua e nel non deben ter cabida as cuestións técnicas, xa que estas serán tratadas en dicionarios de linguaxes especializadas relativas a distintos ámbitos e profesións, pero si reflicte aquilo que un falante cun nivel cultural medio debe coñecer.

Así e todo, autores como Nation (1993a) consideran estas obras que recollen léxico especializado como dicionarios, pero sinala que estes difiren nos criterios para a inclusión de palabras, como son os nomes de plantas e animais, nomes de países, lugares xeográficos, personaxes históricos, de ficción, etc., así como no tratamento de homógrafos, das palabras derivadas con significados próximos á palabra base, que son a miúdo listadas separadamente, ou na inclusión de afixos.

Non hai un criterio uniforme en canto á unidade de tratamento lexicográfico. Os diferentes autores (Alvar, 1993) acóllense a unha variedade de solucións en relación coa inclusión de afixos, de lexías compostas (*de recú, a teor*), modismos, frases e locucións, derivadas, etc. Os autores tamén adoptan distintas solucións en relación cos elementos que figuran como lemas ou forma elixida como representante de tódalas variantes dun vocábulo (habitualmente o infinitivo para os verbos, o singular para os substantivos e o masculino singular para os adxectivos e pronomes).

Os datos que ofrecen os autores de dicionarios sobre o número de entradas non adoitan ser fiables e con frecuencia non sorprende que esaxeren o seu número. Para efectuar un cálculo do número total, como se sinalou no punto anterior, é posible facelo contando cada entrada, por procedementos computerizados ou por mostraxe. Dupuy (1974) e Goulden, Nation e Read (1990) encontraron que os autores do *Webster's Third New International Dictionary* de 1961 informan que contén 450.000, mentres que estes autores calculan arredor de 267.000. O *Shorter Oxford English Dictionary*, ofrece o dato de que contén 163.000, mentres Dolby e Resnikoff (1967) calculan 73.582. Referíndonos a dicionarios de galego, os autores do da RAG (1997) aseguran na introdución que é unha obra dunhas 25.000 entradas, pero, por medio dun conteo manual propio, comprobouse que realmente son 21.364. Parece, por tanto, que no noso ámbito, prodúcese unha tendencia semellante á que se da no anglosaxón.

4.2.2 Elección do dicionario da Real Academia Galega (RAG)

Centrándonos nos dicionarios de lingua galega, López (1990) efectuou unha análise dos máis utilizados (Nós, 1983; Estravís, 1986; Ir Indo, 1986; Xerais, 1986; Galaxia, 1988), e sinala que, aínda que os avances en lexicografía foron importantes, queda aínda un camiño por percorrer. Estes dicionarios recorren fundamentalmente ó portugués e ó castelán, e, en bó número de ocasións, reproducen definicións e acepcións destas linguas, van pasando,

además, duns a outros con escaso rigor. Así, hai palabras idénticas formalmente en galego e catalán ou portugués, pero diferentes semanticamente, e, non obstante, nos dicionarios de galego aparecen como o que a autora denomina *falsos amigos*, é dicir, castelanismos léxicos, semánticos e morfolóxicos ou, se ben con menos frecuencia, como calcos do portugués. Alén disto, segundo esta mesma autora, prodúcese igualmente unha tendencia a incluír cultismos e tecnicismos, fronte ás palabras patrimoniais, autenticamente galegas, así como a excluír absolutamente palabras utilizadas polos diferentes autores da literatura galega, co que se cercena a posibilidade de enriquecemento do léxico.

Un paso previo antes de elixir un dicionario para utilizar nesta tese, consistiu en analizar dicionarios existentes en galego. Deseguido, ofrécese características dalgúns dos máis habituais, como exemplificación das diferenzas existentes entre eles, en canto a cuestións como a información que ofrecen, termos incluídos, número de entradas ou o que se inclúe en cada entrada.

1. Alonso Estravís, I. (1995). *Diccionario Sotelo Blanco da lingua galega*. Santiago: Sotelo Blanco Edicións.

Consta de 1585 páxinas. Non se especifica o número de entradas. Infórmase unicamente de que inclúe léxico científico e popular de uso e que non repecta a normativa da lingua galega. Inclúe os homógrafos numerados en entradas diferentes e prefixos.

En cada entrada, indica a categoría gramatical, acepcións e definicións, con expresións e exemplos, sinónimos e antónimos, nalgúns casos, rematando coa etimoloxía en determinados casos.

2. Ares, M^a. C. et al. (1994). *Diccionario Xerais da Lingua*. Vigo: Xerais (Última edición: X. M^a. Carballeira [Coord.], 2006).

Consta de 993 páxinas e 30.000 palabras, segundo os autores (a edición de 2006, de Carballeira, X.M^a. e outros, contén 37.000 entradas, en 1280 páxinas). Inclúe afixos e os homógrafos etimoloxicamente diferentes en entradas numeradas e diferenciadas, así como termos que, non sendo normativos, os autores consideran lexítimos en lingua galega. Non inclúen derivadas circunstanciais regulares, aumentativos, diminutivos, nin participios sen valor diferencial dos adxectivos.

En cada entrada ofrece a indicación fonética, etimoloxía, categoría gramatical, número da acepción coa definición correspondente, indicacións de uso, exemplos, expresións, frases feitas, así como as variantes, se é o caso, xunto con observacións fonéticas, morfolóxicas, sintácticas, semánticas e ortográficas. Proporciona así mesmo sinónimos e palabras da familia da entrada léxica, moitas das cales non aparecen no lugar que alfabeticamente lle corresponde por ser deducibles da palabra que encabeza a familia. No caso dos sinónimos, remite ós que os autores consideran preferentes.

3. Ares, M^a. C. (1997). *Pequeno diccionario Xerais da Lingua*. Vigo: Xerais.

Contén 16.000 entradas en 777 páxinas (a edición de 2004, coordinada por G. Navaza, contén 18.000 entradas en 626 páxinas). Os criterios que se seguiron para efectuar a redución respecto do *Diccionario Xerais da Lingua* foron os de excluí-las variantes minoritarias de voces máis comúns, as voces de campos específicos, e as de menor frecuencia de uso. En cada entrada inclúense moitas das voces suprimidas correspondentes á súa familia.

4. Boullón, A.I. et al. (1995;5.^a edición). *Diccionario escolar galego*. Vigo: Galaxia.

En 526 páxinas contén arredor de 16.000 entradas. Para a selección os autores utilizaron o *Diccionario normativo galego-castelán* da mesma editorial, coordinado por H. Monteagudo e X. García de 1994 (última edición de 2000), que consta de 40.000, e *Choca. Vocabulario elemental galego*, de X. Xove Ferreiro (Vigo: Galaxia, 1981). Incluíron así mesmo outros termos que se consideraron necesarios nun dicionario escolar básico. O criterio utilizado para a selección das entradas consistiu en eliminar:

- Rexistros científicos e técnicos de escasa presenza na escola.
- Moitos termos creados con prefixación ou sufixación deducibles da palabra base.
- Voces dunha mesma familia léxica, incluíndo só a palabra clave da familia, a partir da que se pode deducir facilmente as demais, a non ser en familias con termos irregulares.
- Variantes de palabras, incluíndo unicamente a máis estendida.

Inclúe os homógrafos etimoloxicamente diferentes numerados en entradas separadas, así como afixos. En cada entrada achega información sobre a categoría gramatical e morfosintáctica, xunto ás acepcións e definicións pertinentes.

5. Feixó, X.G. (1986). *Diccionario da lingua galega*. Vigo: Ir Indo (Última edición, 2004).

Contén 17.000 entradas en 585 páxinas. Na edición de 2004, dirixida por B. Ledo e que se basea na norma do Instituto da Lingua Galega (ILG) e da Real Academia Galega (RAG) de 2003, así como no *Vocabulario Ortográfico da Lingua Galega*, contén máis de 65.000 entradas en 2119 páxinas. Inclúe formas non normativas e estranxeirismos, sinalándoos con (asterisco)* e con indicación da lingua de procedencia, así como neoloxismos e termos científicos e técnicos, acrónimos e abreviaturas (por exemplo, IRPF), ademais de variantes dialectais e de uso. Inclúe os homógrafos numerados en entradas

separadas, así como afixos, e non inclúe palabras rematadas en *-mente* facilmente deducibles.

En cada palabra aparece o nome científico e a materia de coñecemento correspondente, se é o caso, ou a lingua de orixe, categoría gramatical, acepcións con indicacións e campo de uso, locucións, cando procede, e observacións. En canto ós sinónimos, remite a un cando o son en tódalas acepcións e no que se refire ós participios remite ó verbo.

6. García, C. e González, M. (Dir.) (1997). *Diccionario da Real Academia Galega*. Vigo: Edic. Xerais. A Coruña: Galaxia (Última edición, 2000).

En 1241 páxinas, segundo os autores, contén 25.000 entradas, se ben, como xa se indicou anteriormente, mediante unha contabilización manual propia, constátouse que realmente son 21.364. O criterio fundamental de selección foi a frecuencia das palabras. Prescindiron de entradas facilmente deducibles da familia (por exemplo, os rematados en *-mente*). Tampouco aparecen moitas "palabras da rúa", como castelanismos e dialectalismos, por ser este un dicionario da RAG, e cando se recollen márcanse con asterisco (*), remitindo á forma correcta. Inclúen os homógrafos separados en diferentes entradas e afixos.

En cada entrada, achega información sobre a categoría gramatical, definición, acepcións, frases feitas, modismos, sinónimos, antónimos e palabras da familia relacionadas semanticamente, así como, a veces, observacións sobre a pronunciación, ámbito e nivel no que se utiliza habitualmente.

7. Maldonado, C. (Dir.) (1995). *Diccionario didáctico Vieiro*. Ligua Galega. Madrid: SM.

Consta de 867 páxinas e non se especifica o número de entradas. A edición de 2002 consta de 884 páxinas e, aínda que non indica o número de entradas, afirmase que contén 35.000 definicións. Os homógrafos aparecen en entradas diferenciadas e non inclúe sufixos.

En cada entrada achega información sobre a categoría gramatical, definición, fonética con exemplos, tradución a castelán e indicacións de uso, observacións ortográficas, morfolóxicas e gramaticais, sinónimos, antónimos e términos da familia. As palabras non aceptadas pola RAG e ILG aparecen remarcadas con asterisco (*).

8. Pazos, M^a. C. et al. (1993). *Pequeno Dicionario da lingua galega*. A Coruña: RAG.

Contén arredor de 8000 entradas en 478 páxinas. Non informa acerca dos criterios seguidos para a súa redución respecto do *Diccionario da Real Academia Galega* (1997), pero a estrutura é a mesma que a do mencionado dicionario.

9. Xove Ferreiro, X. (Coord.) (1995). *Diccionario da lingua galega*. Vigo: Santillana-Obradoiro.

Consta de 920 páxinas con 20.000 palabras, coas variantes máis usuais. Cando non se corresponden coa normativa, remite á correcta. A reimpresión de 2003, de X. Xove Ferreiro, F. Dubert, E. M. Moscoso e X. C. Sousa, mantén o número de entradas.

Cada entrada vai acompañada dunha referencia á categoría gramatical, acepcións e definición, sinónimos, antónimos, indicacións fonéticas, palabras da familia, frases e expresións feitas, así como observacións. Inclúe os homógrafos numerados en entradas diferentes, así como 1500 entradas incorrectas, nas que remite á correctas ou a outra máis recomendable, pero non inclúe afixos.

Tendo en conta os dicionarios dispoñíbles, a decisión de utiliza-lo dicionario da RAG (García e González, 1997), que, como se sinalou con anterioridade, en 1241 páxinas contén 21.364 entradas, baseouse esencialmente en que, a pesar das diferentes opinións dos especialistas no tema, a Real Academia Galega, xunto co Instituto da Lingua Galega, considérase a autoridade nesta materia, representa o referente e norma para a práctica totalidade das demais obras lexicográficas. Por outra parte, comprende un número aceptable de entradas, fundamentalmente as máis comunmente aceptadas, e exclúe aquelas que non poden ser consideradas estritamente galegas. É de facer notar, non obstante, que, conforme á información facilitada pola RAG e aparecida en *Galicia Hoxe*, con data de 3 de xullo de 2003, nunha próxima edición poden ampliarse o número de entradas ata 40.000.

Con todo, pode considerarse como un inconveniente o feito de que hai palabras que non son aceptadas, pero si utilizadas habitualmente, co cal, aínda que non afecte á estimación do número de palabras coñecidas do dicionario concreto co que se traballa nesta tese, si pode verse reducida a estimación do vocabulario coñecido en xeral.

4.3. Selección das palabras e xeración de pseudopalabras do test

Seguíronse basicamente as contribucións de Goulden, Nation e Read (1990), Nation (1993a) e Suárez, Seisdedos e Meara (1998), expostas no punto 1 deste capítulo, pero non se tiveron en conta en toda a súa extensión as súas indicacións en canto á eliminación de palabras derivadas sinaladas nos puntos 4.1.1 e 4.1.2.

Nunha primeira fase, unha vez contabilizadas manualmente as palabras do dicionario (RAG, 1997), elimináronse os nomes de letras, e as sinaladas con asterisco (*) referidas a términos non aceptados, mentres que os homógrafos contabilizáronse coma unha soa entrada, de forma que a cantidade final foi de 20.624 palabras. Seguidamente, dividiuse o número total das mesmas en bloques de 100, seguindo unha orde alfabética. A continuación, con obxecto de evitar que a selección de vocábulos se inclinase inadecuadamente cara os

ítems que ocupan máis espazo e teñen máis entradas, extraéronse 10 palabras ó azar de cada un dos bloques de 100 palabras establecidos anteriormente, utilizando para isto unha táboa de números aleatorios. Deste modo, obtívose unha mostra de 2062 palabras, que representan o 10% do total de palabras do dicionario.

Nunha segunda fase, para construír dúas baterías de cinco subtests cada unha, seguiuise un procedemento de cinco pasos co obxecto, por un lado, de evita-lo risco de seleccionar unha proporción inadecuada das palabras máis frecuentes do dicionario ou dun determinado grao de dificultade, de xeito que a composición da mostra non representaría convenientemente ás palabras do dicionario (puntos 4.1.2c, 4.1.2g). Por outro lado, a aplicación deste procedemento ten como obxecto corrixi-la probable tendencia dos suxeitos a responder afirmativamente ós ítems sen estar seguros (punto 4.1.3b).

1.º Utilizáronse tres xuíces, docentes do ensino non universitario con ampla experiencia, que impartían ou impartiran lingua galega, alén de ser galegofalantes. Solicitóuselles que puntuasen de 1 a 4 as diferentes palabras da mostra en función do grao de dificultade que, desde o seu punto de vista, representarían para a súa aprendizaxe por parte dos escolares (Anexo I). Deste modo, obtivéronse catro categorías de palabras en función do suposto grao de dificultade, coas seguintes frecuencias de menor a maior dificultade: 1.ª categoría: 601; 2.ª categoría: 579; 3.ª categoría: 571 e 4.ª categoría: 311.

2.º Utilizouse a técnica de mostraxe sen substitución, con afixación proporcional (Sierra Bravo, 1991), co obxecto de que en cada subtest de cada unha das dúas Baterías estivera proporcionalmente representada cada unha das catro categorías mencionadas.

Con este fin, en primeiro lugar, numeráronse as palabras de cada categoría. En segundo lugar, determinouse a proporción de palabras que deberían corresponder a cada categoría, tendo en conta que cada subtest constaría de 40 palabras. Así, cada Batería está

composta por 200 palabras reais, o que representa 1 de cada 100 do dicionario, e todo o Test está composto por 400 palabras, o que representa 1 de cada 50 do dicionario. Esta proporción permite incluír no test un número relativamente alto e manexable de palabras, se ben, como sinala Meara (1993a), sería preferible unha razón máis alta (1 de cada 25, por exemplo), pero o número de ítems en términos absolutos sería excesivamente elevado. Seguindo este procedemento, á 1.^a categoría, que contén 601 das palabras da mostra e representa o 29,15%, correspóndenlle 11,66, é dicir, redondeando 12 palabras; á 2.^a categoría, que consta de 579 e representa o 28,07 %, correspóndenlle 11,21 palabras (redondeando 11); á 3.^a categoría, formada por 571 e representa o 27,69 %, 11,08, é dicir 11 palabras; finalmente, á 4.^a categoría, que contén 311 e que representa o 15,07 %, 6 palabras. En terceiro lugar, seleccionáronse ó azar as palabras de cada categoría para cada subtest, atendendo á proporción que lle correspondía (Táboa 4.2.).

Táboa 4.2.: Número de palabras incluídas en cada subtest

	Categorías			
	1 ^a	2 ^a	3 ^a	4 ^a
Número de Palabras	601	579	571	311
Porcentaxe do Total	29,15	28,07	27,69	15,07
Palabras en cada subtest	12	11	11	6

3.º Seleccionáronse igualmente ó azar unhas 1000 palabras do mesmo dicionario (exactamente 1044), extraendo 5 de cada un dos bloques establecidos na primeira fase. A partir delas xeráronse pseudopalabras, transformándoas mediante a utilización dos criterios que a continuación se expoñen, seguindo a García e Meara (1996):

a) Mante-la lonxitude da palabra.

- b) Mante-la estrutura silábica.
- c) Mante-la acentuación.
- d) Evitar crear termos léxicos que sexan posibles nunha determinada variedade xeográfica ou nun rexistro pouco formal da lingua galega.
- e) Evitar crear palabras que podan existir en español.
- f) Evita-la xeración de pseudopalabras das que se poida recoñecer con facilidade a palabra da que procede (ansinanza, por exemplo).
- g) Evitar producir pseudopalabras que poidan ser identificadas como formas flexionadas de lexemas verbais.
- h) Comprobar que realmente non existen os termos xerados, utilizando para isto o dicionario da RAG (1997) e o da editorial Xerais (1989).

Ademais, solicitóuselle a dous xuíces, profesionais do ensino do galego e con ampla experiencia, que eliminasen aquelas que non terían cabida en galego, que non "soasen" a galego, á vez que as que fosen realmente palabras en galego. Todas elas foron eliminadas e substituídas por outras máis axustadas ó criterio.

4.º De entre as que resultaron aceptadas, seleccionáronse ó azar 20 para cada subtest (Anexo II). Esta proporción de pseudopalabras parece aceptable por unha razón de carácter psicológico e pedagógico: evitar influír negativamente sobre a motivación dos suxeitos, o que podería ocorrer no caso de que se visen obrigados a ter que ignorar ítems nunha alta proporción (Meara, 1994).

5.º Finalmente, numeradas as palabras e pseudopalabras de cada subtest de 1 a 60, mesturáronse aleatoriamente, recolocándoas en canto á orde. Deste modo, quedaron elaboradas dúas probas (baterías), constituídas por cinco subtests cada unha, contando cada

subtest con 40 palabras (12 da 1.^a categoría, 11 da segunda e terceira e 6 da cuarta) e 20 pseudopalabras (Anexo III).

4.4 Sobre a corrección e puntuación

Para o cálculo das puntuacións, neste tipo de tests é preciso ter en conta as seguintes consideracións:

a) En principio, hai dous tipos de respostas correctas (sinalar palabras reais e non sinalar pseudopalabras) e dous tipos de respostas incorrectas (sinalar pseudopalabras e non sinalar palabras reais).

b) Os suxeitos teñen a posibilidade de supoñe-la resposta baseándose nun coñecemento parcial das palabras ou de responder sen estar completamente seguros.

c) Os suxeitos mostran diferentes estilos de resposta (uns presentan unha maior tendencia a sinalar ítems e outros son máis conservadores na emisión de respostas).

Existen catro posibles métodos para calcula-la puntuación: contabiliza-lo número de acertos (ben sen máis, ben contabiliza-lo número de acertos subtraéndolle o número de erros); a corrección das respostas tendo en conta que o suxeito pode emitila porque está seguro ou facéndoo completamente ó azar, e utilizar as medidas *Am* e *ISDT*, elaboradas por Meara, baseadas na “Teoría de Detección de Sinais” e que se explican máis abaixo.

Para a corrección do test construído nesta tese e a obtención das puntuacións mediante as que se consegue a estimación do tamaño de vocabulario coñecido procedeuse seguindo a proposta de Meara e Buxton (1987), Meara (1992, 1994) e Suárez, Seisdedos e Meara (1998). Basicamente consistiu en utiliza-los seguintes pasos en cada unha das cinco subprobos de cada unha das dúas Baterías:

a) Reconto dos ítems sinalados polo suxeito, anotando este dato como Total ou recontando os ítems en branco e restándollos ó total de ítems (60), co que resultan os sinalados.

b) Reconto dos ítems sinalados erroneamente como palabras (pseudopalabras sinaladas como coñecidas), contabilizándoos como Erros.

c) Subtracción do número de Erros ó Total de ítems sinalados, dando como resultado os Acertos.

d) Obtención da media aritmética dos Acertos e dos Erros das cinco subprobas de cada Batería.

e) Consulta desta media na táboa elaborada a partir de principios matemáticos da “Teoría de Detección de Sinais” (Anexo IV), consistente nunha matriz que mostra tódalas posibles combinacións de Acertos e Erros (Suárez, Seisdedos e Meara, 1998, p. 30), da que se expón máis abaixo a súa lóxica. A intersección entre estes dous datos ofrece unha puntuación entre cero e 100 denominada Δm , que sería a estimación do vocabulario coñecido nos subtests e que na práctica se pode interpretar como unha porcentaxe.

Cabe outra posibilidade, consistente en consultar na táboa os valores Δm para cada subtest dunha Batería, e calcula-la media aritmética destes valores (Suárez, Seisdedos e Meara, 1998, p. 13).

f) A continuación, efectúase a estimación do número de palabras coñecidas, multiplicando a puntuación Δm polo número de palabras con que conta a mostra do dicionario utilizada e dividindo por 100.

Como exemplificación destes pasos, expóñense a continuación os datos correspondentes á resposta dun escolar á Batería 1 (Táboa 4.3.).

Táboa 4.3.: Exemplificación de datos utilizados para a estimación do vocabulario

	Subtests					Total	Media	
	1	2	3	4	5			
Ítems sinalados (T)	27	26	26	22	32	133	27	
Erros (E)	2	3	3	2	4	14	3	
Acertos (Total-Erros) (A)	25	23	23	20	28	119	24	
Estimación da porcentaxe de vocabulario (Anexo IV)	42	24	24	24	34	148	30*	28*
Estimación do número de palabras							6187**	5774**

* Puntuación Δm media en función do procedemento seguido ** Número de palabras estimado en función do procedemento seguido para obter a media das puntuacións Δm

Pode comprobarse que obtendo a media de Acertos (24) e Erros (3), e consultando a táboa do Anexo IV, obtense o valor de 28, que representa a estimación do vocabulario en termos de porcentaxe. Multiplicando este valor por 20.624 palabras e dividindo por 100, resulta unha estimación de 5774 palabras.

Seguindo a segunda posibilidade, por medio da consulta das puntuacións de Acertos e Erros na táboa do Anexo IV, obtéñense os valores que aparecen na última fila, de forma que a este escolar estímaselle un vocabulario que vai desde o 24% no segundo, terceiro e cuarto subtest ó 42% no primeiro. Calculando a media aritmética da súa suma (148), obtense unha estimación do 30%. Multiplicando este valor por 20.624 palabras e dividindo por 100, resulta unha estimación de 6187 palabras.

O primeiro procedemento pode resultar o máis aconsellable, pero tamén cabe a posibilidade de obter a media das dúas estimacións, procedemento que se seguiu no presente traballo, de modo que a estimación sería 5980 palabras. A partir deste momento, sempre que se faga referencia á estimación do vocabulario ou á estimación do número de

palabras coñecidas, deberá entenderse que se calculou mediante o método de corrección e estimación que se acaba de explicar (neste exemplo, 5980 palabras).

En canto á lóxica da táboa do Anexo IV, sinalouse que converte o número de Acertos e Erros en estimacións do tamaño do vocabulario. Para a súa elaboración, (Meara, 1991) baseouse en modelos desenvolto por autores como Green e Swets (1966) ou Pastorre e Scheirer (1974), a partir da "Teoría de Detección de Sinais", como xa se comentou en diferentes ocasións.

Calquera combinación de Acertos e Erros pode ser trazada nun eixo de coordenadas, no que os Erros se situarían en x e os Acertos en y (Figura 4.1.). Deste modo, un suxeito que sinalase ó azar ítems sen coñecer ningunha palabra, a posibilidade de marcar tanto unha palabra real como unha pseudopalabra sería próxima ó 50%. Ante esta conduta, probablemente as puntuacións se sitúen na diagonal AC , en tanto que as puntuacións se aproximan a esta diagonal reflicten a decisión de contestar ó azar. Cando a proporción de Acertos é maior que a de Erros, o punto que representa a puntuación situaríase no triángulo ABC . No caso extremo de que os Acertos fosen do 100% e os Erros de 0%, o punto situaríase no vértice B . Nos casos intermedios, a medida que a porcentaxe de Acertos sexa maior que a de Erros, o punto situarase no triángulo ABC . Canto maior sexa o triángulo formado polo punto no que se cruzan Acertos e Erros e a diagonal, maior será o tamaño do vocabulario.

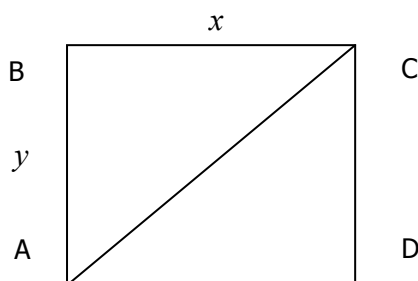


Figura 4.1.: Gráfico da estimación do vocabulario coñecido a partir de Acertos e Erros

Tendo en conta os supostos mencionados máis arriba, Meara utilizou medidas derivadas non paramétricas que considera máis apropiadas para datos relativos ó vocabulario. Estas medidas, que están baseadas na “xeometría da unidade cadrado” (o cadrado no que a área representa unha unidade), son Ag , que é o índice de discriminación ou grao de facilidade con que unha persoa diferenza entre palabras e pseudopalabras, e B' , que é o grao de preferencia dun suxeito na resposta a palabras ou a pseudopalabras.

Non obstante, Ag só é de utilidade para efectuar unha estimación do tamaño do vocabulario cando o número de Erros non é maior que cero. Por isto, Meara realizou unha transformación deste índice Ag , chegando ó índice Δm . A elaboración da táboa que figura no Anexa IV obtense coa utilización da fórmula correspondente ó cálculo deste índice:

$$\Delta m = [(h - f) \times (1 + h - f) / (h \times (1 - f))] - 1$$

onde h é a proporción de Acertos e f a proporción de Erros.

Como aclaración, pode servir un exemplo. Supóñase que a corrección dun test da como resultado que o número de Acertos é de 28 e o de Erros 2, aplicando a fórmula teríamos:

$$[(0,7 - 0,1) (1 + 0,7 - 0,1) / (0,7 \times (1 - 0,1))] - 1 = 0,11 \times 100 = 11$$

valor que, como se pode comprobar, aparece na táboa, cruzando 28 acertos e 2 Erros.

Xa se indicou (capítulo 3, punto 3.2.3) que a precisión da medida Δm é discutible porque, por un lado, cando o número de ítems sinalados é moderado, a puntuación acércase rapidamente a cero, incluso se o patrón de resposta está por riba do nivel do azar, o que deriva nunha subestimación do tamaño do vocabulario. Por outro lado, a precisión de Δm é discutible porque discrepancias no patrón de resposta poden producir importantes diverxencias nas puntuacións, sobre todo cando o número de “Si” a palabras reais é baixo, é dicir, non corrixe o estilo de resposta.

Para rematar este apartado, aínda que non foi posible utiliza-lo nesta tese, debido a que xa se levara a cabo a corrección e tratamento dos datos cando apareceu, é necesario

sinalar que máis recentemente Meara (2002) desenvolveu un novo índice que ten en conta o estilo de resposta e que permite obter cálculos máis precisos do tamaño do vocabulario:

$$ISDT = \frac{1 - 4h(1 - f) - 2(h - f)(1 + h - f)}{4h(1 - f) - (h - f)(1 + h - f)}$$

Con todo, hai aspectos que seguen sendo discutibles como é a asunción de que tódolos ítems presentan o mesmo grao de dificultade, que a tendencia ou estilo de resposta se transfira efectivamente á resposta, ou que unha puntuación non é válida se o número de Acertos é menor que o de Erros.

Capítulo 5

Administración do test e xustificación estatística

5.1 Aplicación piloto

Levouse a cabo unha aplicación do test a modo de estudo piloto co obxecto de comproba-la súa adecuación ó propósito para o que se construíu e o seu funcionamento, resolve-las dúbidas que puideran xurdir ó respecto e introducir, de se-lo caso, as correccións necesarias antes de comeza-la aplicación definitiva das probas. Dada a súa finalidade, nesta fase non se pretendeu traballar cunha mostra que presentara exactamente as características que a que se trataría de utilizar na aplicación definitiva.

Empregouse unha mostra de 43 escolares de sexto de Educación Primaria e 48 de segundo de ESO, pertencentes a centros públicos de zona rural e que nese momento resultou se-la máis accesible. As probas foron aplicadas en dúas sesións con cada grupo, seguindo a pauta que ofrecen Suárez, Seisdedos e Meara (1998): na primeira unha batería de 5 subtests e na segunda os 5 subtests restantes.

Debido a que se traballaba con vocabulario escrito e a que a proba require unha certa precisión e velocidade na lectura que permita ó acceso ó significado das palabras, desde o primeiro momento decidiuse excluír ó alumnado que repetía ou repetira algún curso (1 alumno en sexto de Educación Primaria e 14 en segundo de ESO) e a aquel que presentaba dificultades importantes de lectura e escritura á xuízo do profesorado que impartía as áreas de linguas e ciencias sociais (5 escolares en sexto de Educación Primaria), así como o caso dun inmigrante en sexto que descoñecía a lingua galega, ademais dun alumno con baixa capacidade intelectual en segundo de ESO e con dificultades importantes

en lingua escrita. No caso dos primeiros actuouse así porque a súa idade non se axustaba á que lle corresponde ó alumnado destes cursos, pero tamén porque, aínda que soamente sexa de forma intuitiva, a experiencia mostra que, en boa parte, presentan con frecuencia dificultades, cando menos, na comprensión do que len. No que se refire ós que presentaban importantes dificultades en decodificación da lingua escrita, a decisión adoptouse porque dificilmente poderían resolver a tarefa que se propón adecuadamente. Os datos obtidos son os que aparecen reflectidos na táboa 5.1.

Táboa 5.1.: Estimación do termo medio de palabras coñecidas
(Primeira aplicación piloto)

	6.º de primaria	2.º de ESO
1.ª batería	6042	8835
2.ª batería	5049	9056
Media	5545	8945

Posteriormente, tendo en conta que en sexto se producía unha considerable porcentaxe de escolares con puntuación por debaixo de 1, aínda que Meara (1993a, 1994) indica que nestes casos o máis apropiado é repeti-la aplicación, posto que nesta investigación non sería posible e ante a posibilidade de que se producise debido a unha tendencia a contestar sen respecta-las instrucións, repetiuse a administración con outro grupo de escolares, especificando máis as instrucións e seguindo o mesmo procedemento que na anterior (Anexo V). Utilizáronse 116 escolares de sexto de Educación Primaria, 46 dun centro público do ámbito rural e 70 doutro centro público urbano (tamén neste caso se traballou cos centros en que a accesibilidade foi máis fácil). Unha vez eliminados aqueles que repetían curso (15 escolares), os inmigrantes cun importante descoñecemento da lingua galega e con dificultades na decodificación (5), o alumnado con dificultades en lingua escrita

(3), amais dun caso con baixa capacidade intelectual e dificultades importantes en decodificación, quedaron 92 suxeitos. Nun segundo momento, excluíuse a aquel alumnado co que, unha vez corrixidas as probas, non era posible obter unha estimación do coñecemento do vocabulario, porque sinalaba un excesivo número de pseudopalabras (10), de forma que a mostra definitiva quedou composta por 82 escolares. Neste caso procedeuse deste modo atendendo á suxestión anteriormente citada de Meara (1994), no sentido de que se non é posible volver a repeti-la aplicación das probas, é preferible desestimar os datos destes suxeitos. Os resultados da estimación do vocabulario foron os que figuran na táboa 5.2.

Táboa 5.2.: Estimación do termo medio de palabras coñecidas por escolares de 6.º de Primaria (Segunda aplicación piloto)

	Rural	Urbano	Media
1.ª batería	5736	4258	4997
2.ª batería	6426	5058	5742
Media	6081	4658	

De tódolos modos, continuaba existindo unha porcentaxe de suxeitos para os que parecen manterse as dificultades para respecta-las instrucións ou realiza-las probas con garantías (arredor do 10% na aplicación efectuada por un profesional distinto ó propio investigador). Para tratar de superar esta dificultade, aínda que Meara (1994) sinala que é habitual nunha porcentaxe entre o 5 e o 10%, efectuouse un novo axuste nas instrucións (Anexo VI) e unha variación de formato das probas no seguinte sentido, a fin de facilita-la lectura dos ítems (Anexo VII):

1) Na primeira páxina incluíuse unicamente un espazo para os datos persoais e un exemplo. A versión anterior incluía na primeira páxina espazos para datos persoais e dous subtests.

2) Aumentouse o tamaño da letra e situáronse dous subtests por páxina, agás na última que inclúe un só subtest. Anteriormente incluíanse, como se sinalou, dous subtests na primeira páxina xunto co espazo para datos persoais e tres noutra páxina.

5.2 Selección da mostra. Situación lingüística en Galicia

Antes de pasar á exposición do proceso seguido, é necesario explicar que se adoptou a decisión de incluír información sobre a situación lingüística en Galicia, para xustificala estratificación da mostra do alumnado en distintos agrupamentos, de forma que serviran de base para a formulación de hipóteses e, mediante o contraste das mesmas, poder obter datos sobre a capacidade discriminativa do test e sobre a validez de construto do mesmo. Para a configuración dos grupos atendeuse tanto ó hábitat rural/urbano e costeiro/interior coma ó nivel socioeconómico. Parece sensato pensar que existan diferenzas en canto ó coñecemento do vocabulario en galego entre o alumnado pertencente ós devanditos agrupamentos. Ademais, a información sobre a situación lingüística tamén podería ser utilizada para fundamentar unha distinción de grupos en función da lingua usada no ámbito familiar, que permitise formular unha hipótese ó respecto e igualmente conseguir datos relativos á validez do test.

No que se refire á idade dos escolares da mostra, o ideal houbera sido recoller información sobre o vocabulario coñecido polo alumnado do final de cada ciclo da ensinanza primaria, así como tamén do remate do ensino obrigatorio, é dicir, do segundo ciclo de ESO. Non obstante, isto complicaría enormemente o labor, constituíndo nestes momentos unha tarefa practicamente inviable nunha iniciativa particular como é o caso desta investigación. De tódolos modos, será algo a ter en conta de cara a unha posible ampliación da mesma.

Optouse por estima-lo número de palabras coñecidas polo alumnado de 6.º de Educación Primaria (11-12 anos) e de 2.º de Educación Secundaria Obrigatoria (13-14 anos). A razón pola que nos centramos nestas idades e niveis de escolaridade, reponde ó feito de que son dous momentos, ou cando menos dous deles, realmente importantes na escolaridade. En canto a 6.º de Educación Primaria representa un importante fito no proceso de escolarización, en tanto que se sitúa ó remate dunha etapa, como é a do ensino primario, e ó inicio doutra, coas implicacións que comporta neste caso en canto a esixencias académicas, pero tamén en canto á organización e funcionamento do sistema de escolarización. Polo que se refire a 2.º de Educación Secundaria Obrigatoria, constitúe outro dos fitos do sistema educativo, xa que se encontra ó remate do primeiro ciclo da Educación Secundaria Obrigatoria e o paso ó segundo ciclo, cando menos no momento de aplicación das probas, representaba unha das cuestións que maior polémica xeraban.

Por outra parte, dispoñemos de datos de diferentes investigacións que, como cabería esperar, indican a existencia dun aumento progresivo do tamaño do vocabulario a medida que o fai a idade (por exemplo, Aguado, 1995, 1999; Averril, 1956; Casanova e Rodríguez, 1989; Esteban, 1997; García Hoz, 1946, 1977; Justicia, 1983, 1985a, 1985b, 1995; Rondal, 1982; Suárez, Seisdedos e Meara, 1998). Por este motivo, utilizar dous grupos de distinta idade tamén será de utilidade de cara á obtención da validez de construto do test, en tanto que permita detecta-las diferenzas entre estes grupos de idade.

Coa finalidade de obter unha mostra representativa da poboación destes dous cursos, 6.º de Educación Primaria e 2.º de ESO, e idades, establecéronse dous criterios de estratificación: hábitat e medio socioeconómico. No caso do hábitat, fíxose unha dobre distinción: urbano/rural e costa/interior. Con respecto a esta última distinción, como se explicará máis adiante, fíxose co obxecto de constata-las posibles diferenzas, aínda que nun eido como o que estudamos non dispoñemos de información acerca de máis investigacións

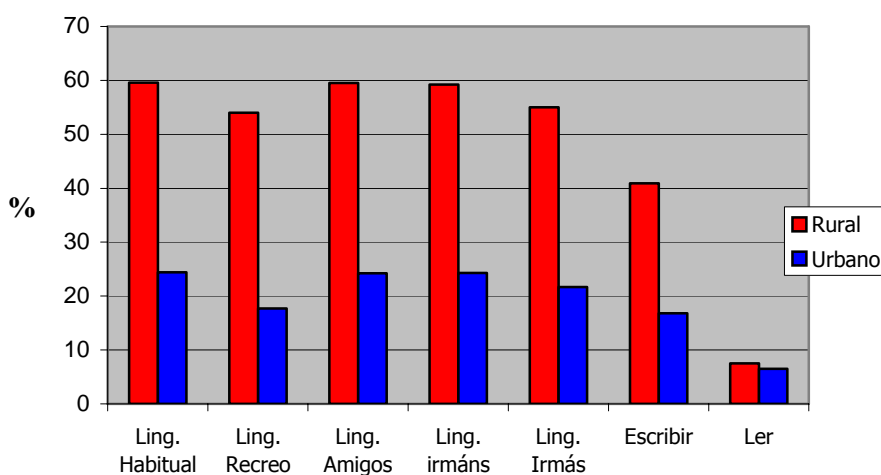
nas que sexa utilizado, a non se-lo traballo de Suárez, Seisdedos e Meara de 1998, no que tiveron en conta esta distinción para a elaboración da mostra galega.

No caso da distinción rural/urbano, partimos do feito de que é unha diferenciación establecida habitualmente na investigación en xeral en Galicia, e máis concretamente naquelas que estudan a utilización e grao de coñecemento da lingua galega. Con todo, os estratos varían nas distintas investigacións. Así, Cajide (1989) utiliza urbano/suburbial/semirural e rural; Fernández (1983) utiliza tanto cidade/vila e aldea coma a dicotomía rural/urbano; Fernández e Rodríguez (1994, 1995, 2000), urbano I e II/ vilas I, II e III/ rural I e II; Pérez Vilariño (1979, 1981) cidade/partido xudicial/vila e aldea; mentres que Rubal e Rodríguez Meira (1987) ou Arza, Rubal e Veiga (1992) agrupan o hábitat en cidade/vila e aldea.

Na presente investigación, tendo en conta as limitacións en que se realiza e o seu carácter experimental, reduciuse o estrato á dicotomía rural/urbano. Neste sentido, considerouse como urbano unicamente as poboacións maiores de 50.000 habitantes, seguindo o criterio da Xunta (Martínez, 2004), e que se corresponden coas cidades existentes na comunidade galega (A Coruña, Ferrol, Lugo, Ourense, Pontevedra, Santiago e Vigo), mentres que se considerou rural o resto. Ó respecto, dispoñemos de investigacións, como as de Justicia (1983, 1985a, 1985b, 1995), que mostran diferencias no tamaño do vocabulario en función desta distinción de hábitat. Por outra parte, os datos do Mapa Sociolingüístico de Galicia (Fernández e Rodríguez, 2000) indican que, en canto ás porcentaxes de persoas que declaran utilizar só ou máis o galego, no sector de poboación entre 16 e 25 anos (o máis próximo ás idades da nosa investigación), existe unha diferenza entre a poboación urbana e a rural de arredor de 35 puntos a favor da rural. A diferenza redúcese a 24 puntos na utilización do galego na escrita e desaparece practicamente en canto á lectura (Táboa 5.3. e Figura 5.1.).

Táboa 5.3. : Utilización da lingua galega entre a poboación de 16 a 25 anos

	Rural %	Urbano %
Lingua habitual	59,6	24,4
Lingua no recreo	54	17,7
Lingua cos amigos	59,5	24,2
Con irmáns homes	59,2	24,3
Con irmás	55	21,7
Para escribir notas	40,9	16,8
Para ler	7,5	6,5

**Figura 5.1.** : Utilización da lingua galega entre a poboación de 16 a 25 anos

Segundo Arza, Rubal e Veiga (1992) a diferenza no uso do galego no ámbito escolar entre a poboación rural e a urbana de 8.º de EXB aproxímase ós 70 puntos porcentuais, se ben neste caso os datos foron proporcionados polo profesorado, non polo propio alumnado.

Respecto da competencia en galego, considerando os datos das seis comarcas que se presentan no informe (Fernández e Rodríguez, 1994), para as idades de 16 a 25 anos, as diferenzas no relativo á porcentaxe de persoas que declaran ter bastante ou moita

competencia sitúanse arredor de 3 puntos para a capacidade de comprensión e en 8 para falar a favor da rural, mentres que as diferenzas están arredor de 15 puntos para as capacidades de ler e para a de escribir a favor da urbana (Táboa 5.4. e Figura 5.2.).

Táboa 5.4. : Dominio da lingua galega da poboación entre 16 e 25 anos

	Rural %	Urbano %
Comprender	98,5	95,7
Falar	82,2	74
Ler	65,6	80,4
Escribir	56,3	71,6

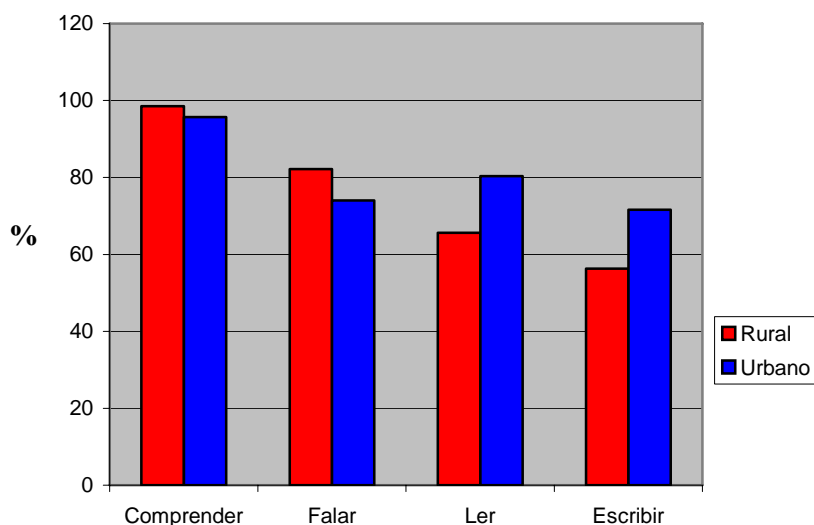


Figura 5.2. : Dominio da lingua galega da poboación entre 16 e 25 anos

No que se refire á poboación xeral de Galicia, que inclúe desde 16 anos a máis de 65 (Táboa 5.5. e Figura 5.3.), as diferenzas entre o rural e o urbano, en canto á porcentaxe de persoas que declaran ser bastante ou moi competentes en galego, tamén diminúen

considerablemente. Así, para a de comprensión é de 3 puntos e para a de falar de 13 a favor da poboación rural; para ler de 9 puntos e para escribir de arredor de 5 puntos a favor da urbana.

Táboa 5.5. : Dominio da lingua galega da poboación xeral

	Rural %	Urbano %
Comprender	97,60	97
Falar	88,05	79
Ler	43	52,5
Escribir	26,90	30

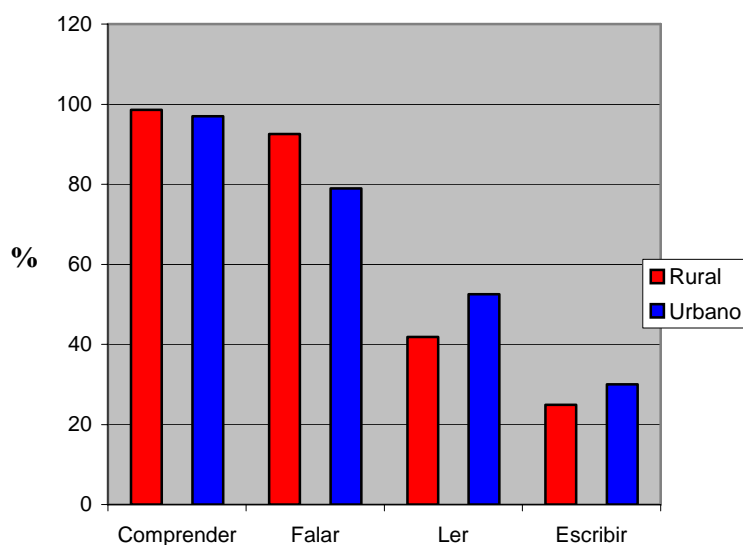


Figura 5.3. : Dominio da lingua galega da poboación xeral

Finalmente, dispoñemos dos datos achegados por Arza, Rubal e Veiga (1992) sobre alumnado de 8.º de EXB, que, se ben foron obtidos a partir da información proporcionada polo profesorado, dalgún modo se corresponderían co remate do actual 1.º ciclo de secundaria, isto é 2.º de ESO (Táboa 5.6. e Figura 5.4.). Segundo estes datos, en opinión

do profesorado, as diferenzas en canto a unha boa comprensión entre o ámbito rural e o urbano son de 7,5 puntos a favor do rural; en expresión, de practicamente 36 puntos; polo que se refire á expresión escrita, os datos contradín ós do Mapa Sociolingüístico de Galicia (MSG), xa que a diferenza é 26,9 puntos a favor do rural.

Táboa 5.6.: Dominio da lingua galega polo alumnado de 8.º de EXB

	Rural %	Cidade %
Comprensión oral	88,80	81,32
Expresión oral	72,50	36,50
Expresión escrita	62,38	35,48 %

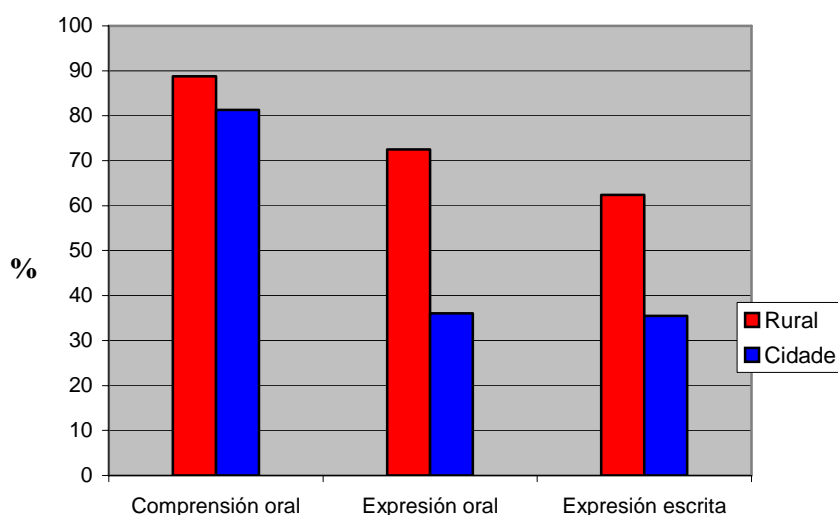


Figura 5.4.: Dominio da lingua galega polo alumnado de 8.º de EXB

En xeral, polo tanto, parece que podemos considerar que oralmente a poboación do ámbito rural utiliza máis e síntese máis competente en lingua galega que a poboación urbana. Con todo, no que respecta á escrita os datos varían. Así, para a poboación entre 16 e 25 anos, mantense a orientación das diferenzas en canto á súa utilización para a escritura de notas, mentres que na lectura iguálanse practicamente as porcentaxes. Pola súa parte, segundo a investigación de Arza, Rubal e Veiga (1992) o alumnado de 8.º de EXB do rural

aparece como máis competente en galego. Pero, en base á información achegada por Fernández e Rodríguez (1994) sobre a percepción da propia competencia, os datos invértense na súa tendencia no referente á poboación de 16 a 25 anos, xa que a urbana declárase máis competente que a rural, tanto na capacidade para escribir como na de ler en galego. Algo semellante a isto último ocorre coa poboación xeral (maior grao de competencia da poboación urbana), aínda que no relativo á competencia para ler a diferenza redúcese considerablemente (a 3 puntos porcentuais).

En síntese e en termos xenéricos, podemos concluír que oralmente a poboación rural utiliza máis e declárase máis competente en galego, á vez que o utiliza máis na escrita, mentres que a urbana declárase máis competente en galego na escrita.

O segundo criterio de estratificación dentro do hábitat (costa/interior), utilizouse en tanto que é unha diferenciación empregada en determinadas investigacións no ámbito da comunidade galega, debido a que existen diferenzas en canto á actividade económica e o modo de vida en xeral, derivada da mesma.

Como xa se sinalou, apenas dispoñemos de información sobre investigacións que aborden temáticas semellantes á deste traballo e que utilicen esta distinción. Non obstante, adoptouse a decisión de incluíla porque, como se desprende da análise dos datos que se comentan a continuación e coas pertinentes reservas, tendo en conta que a fonte utilizada non establece explicitamente a distinción, pódese pensar que existen certas diferenzas no referente á utilización do galego.

Non é posible obter datos de todo Galicia, xa que para o agrupamento de concellos nos sectores utilizados para a elaboración do Mapa Sociolingüístico de Galicia (MSG) algúns deles están constituídos por concellos costeiros e de interior. De tódolos xeitos, a partir dos datos de Fernández e Rodríguez (2000), pódese facer unha aproximación ás posibles diferenzas en canto á utilización do galego das zonas de costa respecto das de interior, se se traballa cos datos de comarcas constituídas unicamente por concellos de costa ou de interior

(que son 16 dos 34 que compoñen a mostra total). Deste modo (Táboa 5.7. e Figura 5.5.), traballando con sete sectores de costa e nove de interior, para o rango de 16 a 25 anos, a diferenza a favor da zona de interior na utilización do galego falado é de entre 23 e 28 puntos porcentuais, en canto á porcentaxe de persoas que declaran usar só ou máis o galego. Respecto á utilización na escrita, a diferenza sitúase ó redor dos 19 puntos, diferenza que baixa a menos de un punto na lectura. Aténdose a estes datos, parece que nas zonas de interior utilízase en maior medida o galego que nas zonas de costa.

Táboa 5.7.: Utilización da lingua galega entre a poboación de 16 a 25 anos

	Costa %	Interior %
Lingua habitual	30,25	59,8
Lingua no recreo	27,30	50,8
Lingua cos amigos	31,40	59,7
Con irmáns homes	28,30	56,2
Con irmás	30,90	58,9
Para escribir notas	21,10	40,4
Para ler	6,60	7,3

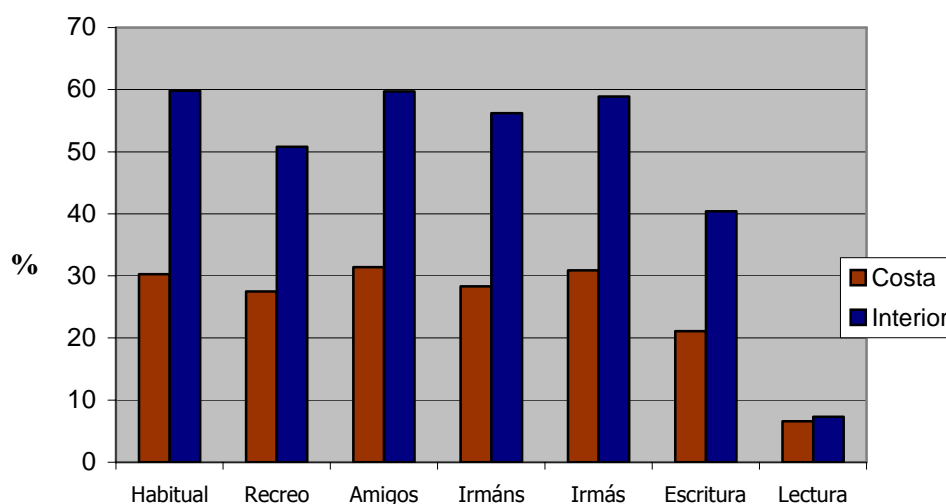


Figura 5.5.: Utilización da lingua galega entre a poboación de 16 a 25 anos

O segundo criterio xeral de estratificación (medio socioeconómico), estableceuse debido a que é unha das diferenciacións que se utilizan en investigacións sobre aspectos como os abordados nesta ou semellantes, en base ás diferenzas existentes en canto ós indicadores habituais de nivel socioeconómico (estudos, profesións, ingresos, consumo, medio residencial familiar, etc.). Entre as investigacións a que se fai referencia están, por exemplo: Fernández (1983), Fernández e Rodríguez (1994, 1995, 2000), García Hoz (1946, 1977), Justicia (1983, 1985a, 1985b, 1995), Pérez Vilariño (1979) ou Suárez, Seisdedos e Meara (1998).

Con todo, presentouse unha dificultade de carácter práctico para selecciona-la mostra en base a este criterio utilizando os indicadores habituais, xa que para establece-la mostra atendendo a estes indicadores, sería preciso recoller esta información sobre os suxeitos con anterioridade, e posteriormente elabora-la mostra, o cal complicaría e convertería en impracticable o proceso nestas circunstancias. Por esta razón, establecéronse os niveis socioeconómicos e culturais en base ó tipo de centro en función da súa titularidade, xa que nos privados, polas súas demandas económicas, asiste alumnado pertencente a un nivel alto ou medio-alto, mentres que, en termos xenéricos, nos concertados e nos públicos do centro de vilas e cidades a poboación escolar habitualmente procede do nivel medio/medio-baixo; e, finalmente, nos públicos de extrarradio e rurais a poboación maioritaria correspóndese co nivel medio-baixo e baixo. Deste modo, identificámo-la clase social coa titularidade do centro en combinación coa localización (García Hoz, 1946, 1977; Suárez, Seisdedos e Meara, 1998).

No relativo a esta variable, existen investigacións españolas que poñen de manifesto diferenzas no tamaño do vocabulario en función da clase social de pertenza (García Hoz, 1946, 1977; Justicia, 1983, 1985a, 1985b, 1995; Suárez, Seisdedos e Meara, 1998). Por outra parte (Táboa 5.8. e Figura 5.6.), os datos do Mapa Sociolingüístico de Galicia indican que dentro do sector de poboación entre 16 e 25 anos (o máis próximo ás idades desta investigación) existen diferenzas na utilización do galego falado entre os niveis

socioeconómicos baixo/medio-baixo e medio-alto de entre 27 e 35 puntos porcentuais a favor do baixo e medio-baixo. A porcentaxe de utilización do galego é de entre 7 e 14 puntos superior nos niveis baixo/medio-baixo en relación coa utilización deste idioma no nivel socioeconómico medio. Polo que respecta á escrita, a diferenza redúcese a uns 18 puntos (respecto da clase media máis de 13 puntos), e na lectura a 1 punto aproximadamente (respecto da clase media non existe diferenza).

Táboa 5.8. : Utilización da lingua galega entre a poboación de 16 a 25 anos

	Baixa/media baixa %	Media %	Media alta %
Lingua habitual	60,90	39,8	28,3
Lingua no recreo	54,10	34,7	27
Lengua cos amigos	60,30	40,3	28,4
Con irmáns homes	59,40	38,5	23,8
Con irmás	54	36,1	21,9
Para escribir notas	40,95	27,8	23,4
Para ler	7	6,9	8,4

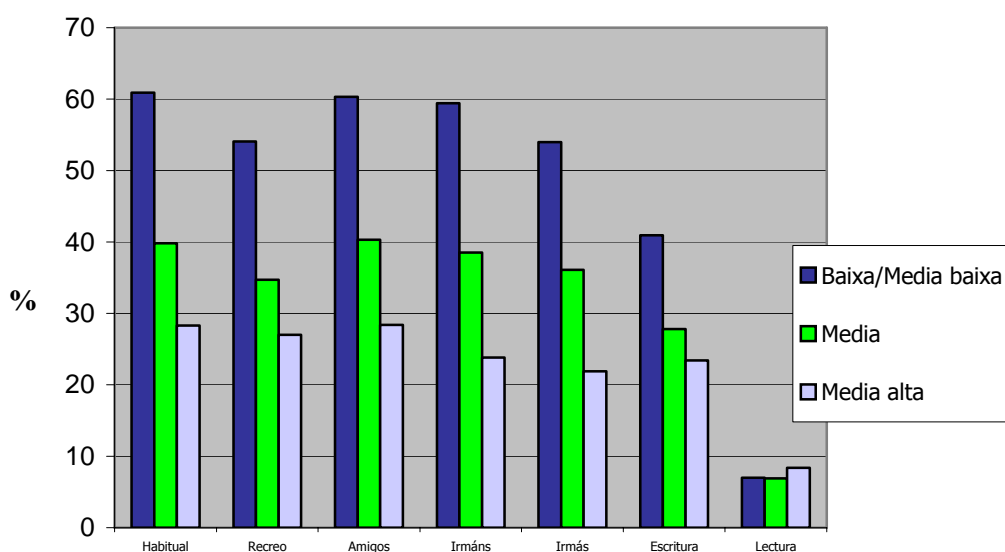


Figura 5.6. : Utilización da lingua galega entre a poboación de 16 a 25 anos

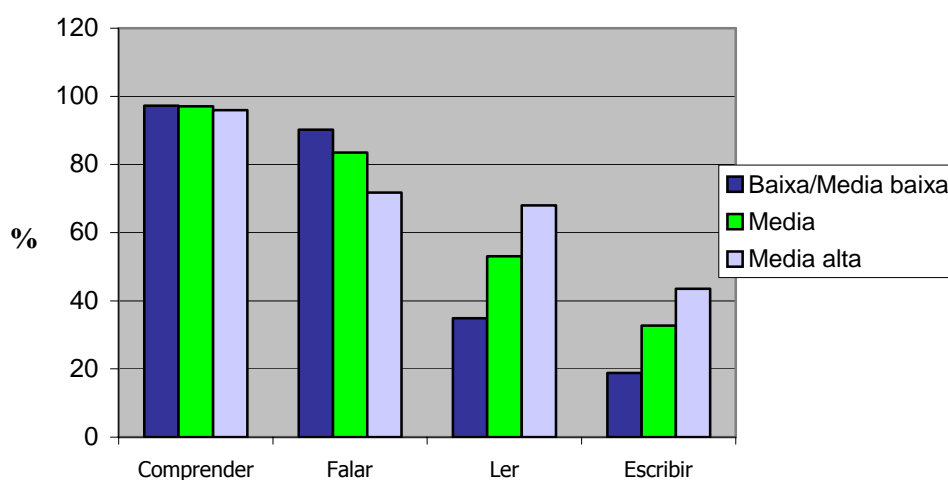
A tendencia segue a mesma dirección segundo os datos aportados por Pérez Vilariño (1979) para a poboación urbana e tamén segundo os de Fernández (1983), en relación co alumnado de EXB.

Se se atende á distinción baseada na titularidade do centro educativo, os datos achegados por Arza, Rubal e Veiga (1992) mostran unha diferenza na utilización do galego no medio escolar de máis de 36 puntos porcentuais a favor do alumnado de 8.º de EXB de centros públicos. Conforme os datos acadados por Santos, Cajide, Lorenzo e Porto (2002) con alumnado de 6.º de Educación Primaria, as porcentaxes de utilización do galego nos contextos familiar e escolar é claramente superior no caso do alumnado que asiste a centros públicos, diferenzas que se sitúan por encima do 40% en tódolos casos.

En canto á competencia en galego (Táboa 5.9. e Figura 5.7.), no que se refire á poboación xeral de Galicia (Fernández e Rodríguez, 1994) a diferenza en comprensión é de un punto aproximadamente a favor dos niveis baixo e medio-baixo respecto do nivel medio-alto, no relativo á porcentaxe de persoas que declaran ser bastante ou moi competentes en galego (respecto á clase media non existe diferenza). No que se refire á competencia para falar, á diferenza é de 18 puntos porcentuais a favor da clase baixa e media-baixa, mentres que é de 7 puntos respecto á clase media. Non obstante, no referente á competencia en lectura, a diferenza é a favor da media-alta de 33 puntos (de 15 respecto da clase media) e na de escribir esta diferenza é de 25 puntos (11 respecto da clase media).

Táboa 5.9.: Dominio da lingua galega da poboación xeral

	Baixa/media baixa %	Media %	Media alta %
Comprender	97,2	97,1	95,9
Falar	90,2	83,5	71,8
Ler	34,9	53	68
Escribir	18,8	32,7	43,5

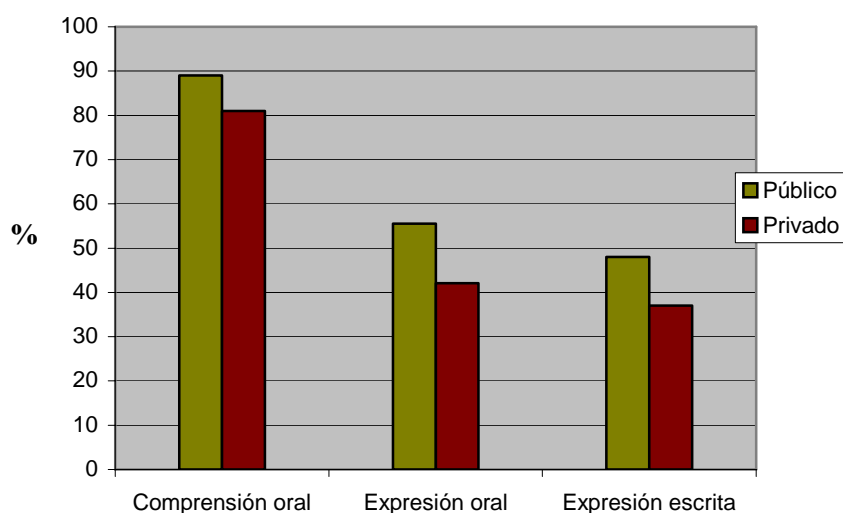
**Figura 5.7.:** Dominio da lingua galega da poboación xeral

En termos xerais, parece que as clases baixa e media-baixa utilizan máis o galego que a media-alta, tanto para falar como para escribir, pero no relativo á lectura as diferenzas son a favor da media-alta, se ben moi reducidas. Tamén parece que a poboación pertencente ás clases baixa e media-baixa sintese máis competente na lingua oral en galego que a media-alta, pero na escrita os datos invértense.

Se se utiliza a diferenciación baseada na titularidade do centro escolar, a investigación de Arza, Rubal e Veiga (1992) mostra diferenzas a favor dos públicos segundo a opinión do profesorado, en canto á capacidade do alumnado de 8.º de EXB para entender, expresarse oralmente e escribir ben en galego (Táboa 5.10. e Figura 5.8.).

Táboa 5.10.: Dominio da lingua galega do alumnado de 8.º de EXB

	Público %	Privado %
Comprensión oral	89,14	81,03
Expresión oral	55,49	42,10
Expresión escrita	48,34	36,90

**Figura 5. 8.:** Dominio da lingua galega do alumnado de 8.º de EXB

Finalmente, utilizando a información que proporciona o Mapa Sociolingüístico de Galicia (MSG), tamén se tratou de diferenza-la mostra en función do que se denomina aquí lingua habitual ou usual no medio familiar. Intentouse conseguir este dato mediante unha pregunta incluída na portada do protocolo do Test, situada entre outros datos de carácter persoal: “Coa miña nai e meu pai falo en galego ☐ castelán ☐ nas dúas ☐”. En principio, a formulación da pregunta ocasiona un certo equívoco, xa que se podería pensar que proporciona información relativa á lingua habitual na familia, pero tamén que a información que subministra é a relativa á lingua inicial. Tamén pode resultar equívoco o non distinguir entre a fala co pai e coa nai. Tendo en conta este feito e que o MSG proporciona información diferenciada sobre a lingua inicial e sobre a habitual, preferiuse comenta-la información de

forma diferenciada sobre a lingua habitual e inicial, sen aborda-la distinción no que se refire á lingua falada co pai e coa nai. En base a estes datos, cabe sinalar que as persoas que teñen o galego como lingua inicial (lingua na que aprenderon a falar) superan na capacidade de falar en galego en case 40 puntos porcentuais ás que teñen o castelán como lingua inicial e en 10,6 puntos ás que teñen as dúas. No que se refire á lectura e escritura, a situación é a inversa, xa que son as persoas que aprenderon a falar en castelán as que din ter moito ou bastante dominio en galego nunha maior porcentaxe que as que aprenderon a falar en galego (Táboa 5.11. e Figura 5.9.).

Táboa 5.11.: Dominio da lingua galega da poboación xeral segundo a lingua inicial

	Galego %	Castelán %	As dúas %
Comprender	99,3	91,7	97,3
Falar	97,9	58,7	87,3
Ler	40,1	56,1	53,8
Escribir	22,2	34,4	36,3

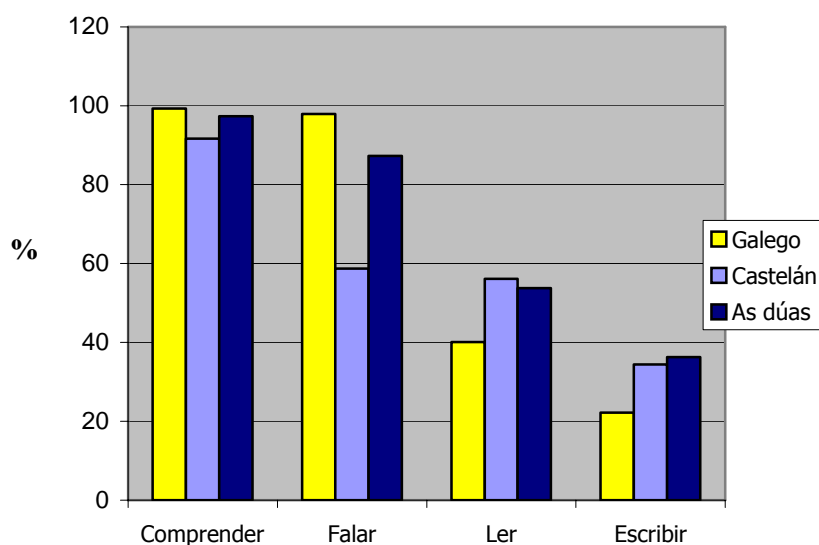


Figura 5.9.: Dominio da lingua galega da poboación xeral segundo a lingua inicial

Para as idades entre 16 e 25 anos, o mesmo informe unicamente ofrece datos relacionados coa lectura para seis comarcas, pero resulta interesante observar que nestas idades é maior a porcentaxe de persoas galegofalantes cun bo ou bastante dominio (Táboa 5.12. e Figura 5.10.).

Táboa 5.12.: Dominio da lectura en galego segundo a lingua inicial entre a poboación entre 16 e 25 anos

Lingua inicial	Dominio da lectura en galego (%)
Galego	78,63 %
Castelán	70,55 %
As dúas	77,70 %

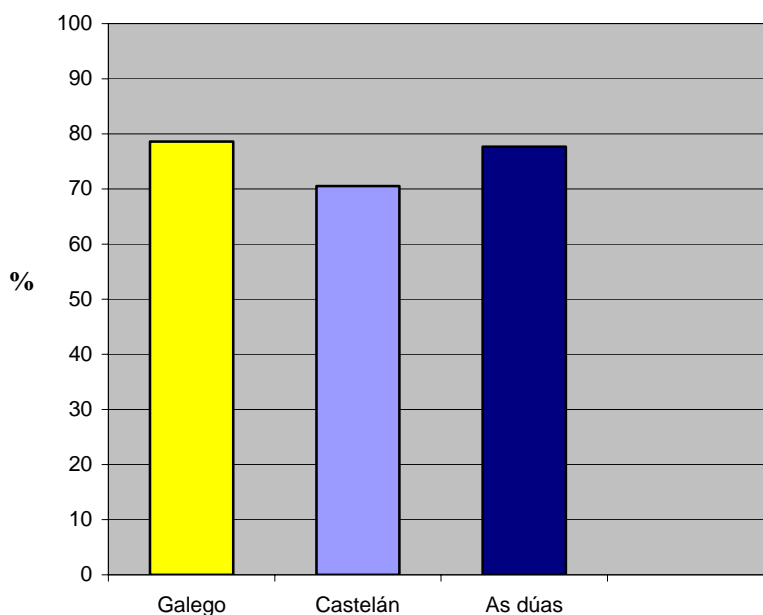


Figura 5.10. Dominio da lectura en galego segundo a lingua inicial entre a poboación entre 16 e 25 anos

En canto á relación da lingua habitual co dominio do galego na poboación xeral, os datos indican que se produce unha situación semellante á da lingua inicial: as persoas que utilizan só ou máis o galego din ser máis competentes en galego no plano oral, mentres que as que falan só ou máis en castelán din ser máis competentes en galego na lectura e escritura que os galegofalantes (Táboa 5.13. e Figura 5.11).

Táboa 5.13.: Dominio da lingua galega segundo a lingua habitual da poboación xeral

	Só ou máis en galego %	Só ou máis en castelán %
Comprender	99,25	89,65
Falar	98,30	51
Ler	43,15	52,15
Escribir	24,60	31,55

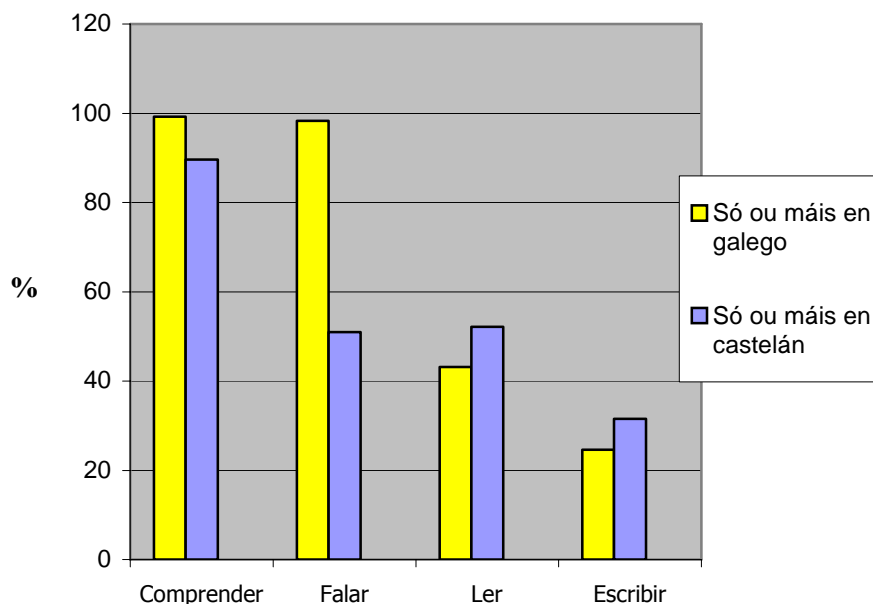


Figura 5.11.: Dominio da lingua galega segundo a lingua habitual da poboación xeral.

Resumidamente, aínda que os datos das diferentes investigacións non concorden en toda a súa extensión, en termos globais parece que se pode concluí-lo seguinte:

a) A poboación rural utiliza a lingua galega na fala en maior medida que a urbana e considérase con maior competencia nela, ó tempo que tamén utiliza esta lingua máis frecuentemente na escrita. Non obstante, é a poboación urbana a que se considera con maior grao de competencia na escrita en galego.

b) A poboación de zonas de interior utiliza máis o galego que a costeira, tanto oralmente coma na escrita.

c) A poboación de nivel socioeconómico baixo e medio-baixo utiliza o galego máis que a procedente dun nivel medio e medio-alto, agás para a lectura. Esta mesma tendencia prodúcese comparando alumnado de centros públicos e privados, a favor do de centros públicos. Polo que se refire ó grao de competencia, as persoas de nivel socioeconómico baixo e medio-baixo considéranse máis competentes en galego que as que proceden do nivel medio e medio-alto en relación á lingua oral, pero en lectura e escritura sucede ó contrario. Se se compara a competencia que en opinión do profesorado posúe o alumnado segundo a titularidade do centro, o de centros públicos presenta maior competencia en galego en tódolos planos da lingua.

d) En termos globais, as persoas que teñen o galego como lingua inicial ou que a utilizan máis habitualmente defínense como máis competentes no plano oral en galego, mentres que as persoas monolingües en castelán ou que utilizan este idioma máis habitualmente defínense como máis competentes que os galegofalantes na lectura e escritura en galego. Segundo os datos do MSG, a única excepción a esta tendencia represéntana as persoas menores de 26 anos en relación coa lectura, xa que son os galegofalantes os que se definen como máis competentes en galego.

Tendo en conta esta información, parece que podería ser adecuado utiliza-lo contraste das posibles diferenzas no tamaño do vocabulario entre os escolares que

pertencen ós distintos agrupamentos como un medio para obter datos relativos á validez de construto do test.

Polo que respecta á selección da mostra, Butcher (1973, pp. 26-28) propón a título de suxestión unha secuencia no que se refire ás diferentes posibilidades para levar a cabo o proceso:

1) Definir clara e inequivocamente a poboación sobre a que se intentan xeneraliza-los resultados da investigación.

2) Estima-lo tamaño da mostra coa que traballar, tendo en conta: a) os subgrupos dentro dela e o tamaño dos mesmos, b) o grao de estabilidade estatística da mostra e dos subgrupos, c) as posibilidades en canto a tempo e dificultade para a recollida e análise de datos.

3) Considera-la posibilidade de utilizar unha mostraxe aleatoria simple ou o seu equivalente, unha mostraxe aleatoria sistemática, aínda que no ámbito educativo non resulta facilmente realizable.

4) Incluso se é posible unha mostraxe aleatoria simple, é probable que o uso dunha estratificación incrementa substancialmente a conveniencia ou precisión da mostraxe. Para tentar conseguir maior precisión, é necesario avalía-la información sobre a asociación entre estratos "naturais", como sexo, idade ou tipo de escola, e variables psicolóxicas ou de outro tipo relacionadas coa investigación. Efectuada esta valoración, é aconsellable elixir dous ou tres máis relacionados coa principal variable de investigación, xa que máis estratos complicarán o traballo, en canto ás proporcións a incluír en cada cela.

5) Se non é posible nin unha mostraxe aleatoria simple nin estratificada, intentar preservar-lo principio de selección aleatoria escollendo os grupos aleatoriamente, de forma que poden darse diferentes posibilidades. Pode ser mellor estratifica-lo tipo de escola e escoller escolas de cada tipo aleatoriamente, ou coller unha mostra aleatoria simple de escolas, logo de aulas ou nenos dentro das escolas, ou adoptar outro deseño. É conveniente

ter en conta as consideracións expostas en 2) e a variabilidade entre escolas ou outros agrupamentos respecto da variable a estudar (canto maior sexa, menor será a precisión do método de mostraxe estratificada). Se as condicións son semellantes noutros aspectos, varios estratos máis pequenos funcionarán mellor que poucos máis grandes.

6) Utilizando calquera dos métodos sinalados, será posible estima-la precisión dos resultados estatísticos obtidos. Se se calcula a media ou unha proporción dentro da mostra, os datos serán válidos para determina-lo erro típico desa media ou proporción e establece-los límites de confianza correspondentes ós estatísticos da poboación. Pero as fórmulas que ofrecen os textos habitualmente só son aplicables ás mostraxes aleatorias simples. Canto máis complicado sexa o deseño, máis o será a fórmula para calcula-lo erro estándar.

7) Se non é práctico usa-la selección aleatoria en ningunha etapa, intentar aplica-los principios xa mencionados, que afectan á estratificación e mostraxe de grupos completos. Nada do dito significa, por exemplo, que a estratificación sen selección ó azar careza de valor. Poden facerse estimacións axustadas da poboación con mostras non aleatorias (enquisas de opinión, por exemplo). A diferenza crucial é que non é posible facer unha estimación teoricamente satisfactoria do erro de mostraxe que sempre está presente.

8) Incluso unha mostra "accidental" ou unha elixida porque é a única posible, pode proporcionar información valiosa. Isto é máis certo en áreas pouco exploradas e con estudos clínicos intensivos de poucos suxeitos. Con todo, o esencial é que o investigador explicita a fortaleza ou debilidade do deseño da mostraxe. O valor das conclusións dependerán enormemente da análise de ata qué punto foron satisfeitas as condicións para unha xeneralización científica.

Na presente investigación, en base ó exposto anteriormente, nun primeiro momento, procedeuse atendendo ós seguintes pasos para a extracción da mostra:

Primeiramente, partiuse dos datos facilitados pola Consellería de Educación e Ordenación Universitaria da Xunta de Galicia (1999), consonte ós que no curso 1998/99

estaban escolarizados 28.800 alumnos en sexto de Educación Primaria e 36.382 en segundo de Educación Secundaria Obrigatoria.

Tendo en conta estes datos, nun primeiro paso, efectuouse un cálculo para o tamaño da mostra total a empregar, cunha marxe de erro de mostraxe do 5%, co fin de coñecer o tamaño da mostra ideal, como unha referencia básica á hora de xulgar a mostra que fose posible empregar na práctica. Para elo, utilizouse a fórmula para poboacións finitas (Ferraces e Andrade, s.d.; Sierra Bravo, 1991):

$$n = \frac{Z^2 \cdot p \cdot q \cdot N}{E^2 (N-1) + Z^2 \cdot p \cdot q} = \frac{2^2 \cdot 50 \cdot 50 \cdot 65182}{5^2 \cdot (65182 - 1) + 2^2 \cdot 50 \cdot 50} = 397 \approx 400$$

Na fórmula, Z correspóndese co nivel de confianza elixido, que neste caso é o 95%, co que o nivel de significación é 0,05 e Z é 1,96, que se aproxima a 2; p e q son as proporcións; E representa o erro que se asume, N o tamaño da poboación e n o tamaño da mostra.

Nun segundo paso, estableceuse proporcionalmente o número de escolares que lle corresponden á poboación de 6.º de Educación Primaria ($n=177$) e á de 2.º de ESO ($n=223$), así como os que lle corresponden a cada estrato das outras variables elixidas. Nas Táboas 5.14. e 5.15. figura o número de escolares que lle corresponden a cada agrupamento e, entre paréntese, a proporción dos mesmos que lle correspondería. Na Táboa 5.16. figura a cantidade de escolares que lle corresponderían proporcionalmente a cada grupo e a porcentaxe que representan do total da mostra.

Táboa 5.14.: Poboación e mostra de 6.º de Educación Primaria

		Rural			Urbano			Totais
		Públ.	Conc.	Priv.	Públ.	Conc.	Priv.	
Costa	Coruña	3076 (19)	645 (4)		1329 (8)	1716 (11)	141 (1)	(94) (53%)
	Lugo	501 (3)	141 (1)	---	---	---	---	
	Ourense	---	---	---	---	---	---	
	Pontevedra	2842 (17)	609 (4)		1851 (11)	2201 (14)	115 (1)	
Interior	Coruña	3317 (16)	201 (1)		400 (4)	588 (5)	152 (2)	(83) (47%)
	Lugo	1737 (11)	162 (1)	5 (--)	509 (3)	456 (3)	27 (--)	
	Ourense	1516 (9)	276 (2)	---	609 (4)	716 (4)	13 (--)	
	Pontevedra	2556 (16)	325 (2)	36 (--)	---	---	---	
Totais		(86%)(91)	(14%) (15)	0	(42%) (30)	(52%) (37)	(6%) (4)	177
		(60%) (106)			(40%) (71)			

Táboa 5.15.: Poboación e mostra de 2.º de ESO

		Rural			Urbano			Totais
		Públ.	Conc.	Priv.	Públ.	Conc.	Priv.	
Costa	Coruña	4205 (26)	668 (4)	---	2092 (13)	1863 (11)	170 (1)	(117) (52%)
	Lugo	671 (4)	143 (1)	---	---	---	---	
	Ourense	---	---	---	---	---	---	
	Pontevedra	3483 (22)	646 (4)	60 (--)	2456 (15)	2498 (15)	140 (1)	
Interior	Coruña	4192 (26)	186 (1)	---	729 (4)	542 (4)	216 (1)	(106) (48%)
	Lugo	2216 (14)	202 (1)	8 (--)	710 (4)	444 (3)	42 (--)	
	Ourense	1791 (11)	311 (2)	30 (--)	1029 (6)	847 (6)	39 (--)	
	Pontevedra	3359 (21)	349 (2)	45 (--)	---	---	---	
Totais		(89%) (124)	(11%) (15)	0	(50%) (42)	(46%) (39)	(4%)(3)	223
		(62%) (139)			(38%) (84)			

Táboa 5.16: Resumo das mostras calculadas para 6.º de Educación Primaria e 2.º de ESO

		6º. Educación Primaria	2º. ESO	Totais
Hábitat 1	Rural	106 (60%)	139 (62%)	245 (61%)
	Urbano	71 (40%)	84 (38%)	155 (39%)
	Total	177 (44%)	223 (56%)	400
Hábitat 2	Costa	94 (53%)	117 (52%)	211 (53%)
	Interior	83 (47%)	106 (48%)	189 (47%)
	Total	177 (44%)	223 (56%)	400
Titularidade do centro	Público	121 (68%)	166 (75%)	287 (72%)
	Concertado	52 (29%)	54 (24%)	106 (26%)
	Privado	4 (2%)	3 (1%)	7 (2%)
	Total	177 (44%)	223 (56%)	400

Para seguir coa terceira fase, xurdiu unha dificultade á hora de establece-la categoría de clase social, como xa se sinalou. Optouse, como xa se indicou tamén, por establece-la equivalencia titularidade do centro/clase social en combinación co hábitat (centro de vila ou cidade/extrarradio/rural).

Para poder controlar con rigor esta variable sería preciso conseguir datos sobre a súa procedencia social, en cada centro e de cada escolar, para posteriormente extrae-la mostra que lle correspondería a cada estrato. Outra posibilidade sería ir agrupando centros públicos de zonas do centro das cidades con concertados. De tódolos modos, calquera das posibilidades gravaría enormemente o proceso de selección das mostras.

Por outra parte, como se pode apreciar nas Táboas 5.14. e 5.15., procedendo deste modo prodúcense situacións en canto á viabilidade/aplicabilidade, dificilmente asumibles. Así, por exemplo, en relación coa categoría clase social, non tería cabida alumnado de

ámbito rural costeiro na privada ou unicamente dous alumnos na privada urbana de costa e outros dous na privada urbana de interior. Os datos da mostra tampouco permitirían efectuar comparacións entre o alumnado de 6.º de Educación Primaria pertencente ós distintos agrupamentos establecidos (rural/urbano, costa/interior, etc.) nin entre o alumnado de 2.º de ESO correspondente ós devanditos agrupamentos, xa que para poder contrastar estas diferenzas sería preciso traballar cunha mostra moito maior. Por outra parte, nesta tese, máis que levar a cabo un estudo representativo de toda a poboación, trátase de elaborar un instrumento que permite estimar cantas palabras coñece un determinado suxeito do dicionario da RAG, se ben tamén se tratará de elaborar uns baremos, aínda que de momento unicamente se faga en función do nivel de escolarización (6.º de Educación Primaria e 2.º de ESO) e como unha proposta inicial que sirva de base para posteriores traballos.

Neste sentido, fundamentalmente interézanos demostrar que as probas elaboradas serven para obter estimacións fiables do vocabulario galego que posúe un determinado escolar. Para isto, é preciso demostrar que as probas son fiables, que permiten efectuar a mesma estimación, con marxes de erro coñecidas e asumibles, se se lle aplican en diferentes ocasións á mesma persoa, e válidas, é dicir, que realmente sexan de utilidade para medir o tamaño do vocabulario.

Por estas razóns, aínda que a idea directriz era acercarse o máis posible a unha mostra proporcional estratificada aleatoria, tendo en conta as limitacións dunha iniciativa privada como é a presente investigación e a perda de suxeitos na aplicación piloto, optouse por administralle o test a unha mostra ampla de suxeitos que cumpriran os requisitos en relación coas características definidas para os diferentes estratos establecidos. Para isto, utilizaríanse os centros nos que fose posible a aplicación.

5.3 Administración definitiva. Mostra resultante

Nesta fase, dadas as características da investigación, que se desenvolveu como unha iniciativa particular, para a elección dos centros e aplicación das probas, utilizáronse fundamentalmente contactos persoais coas direccións e xefaturas de Departamentos de Orientación dos centros educativos, agás no caso dos centros de titularidade privada, en que se solicitou a intervención das respectivas inspeccións educativas de zona, que facilitaron o trámite.

Non se seleccionaron centros concertados, debido a que, en boa parte, o tipo de alumnado non se diferencia esencialmente do dos públicos. Tampouco se intentou seleccionar centros privados de zona rural, debido a que a súa implantación nestas zonas socioxeográficas é moi limitada.

En principio, aínda que, como se explicou no apartado anterior, inicialmente pretendíase traballar cunha mostra proporcional estratificada, finalmente non se conseguiu, debido a que unicamente se puido aplica-lo test naqueles centros que xeograficamente resultaban máis accesibles e que voluntariamente accederon. En consecuencia, a mostra resultante poderíase cualificar como unha combinación de “dispoñible” ou “viable” e “intencional” (Azorín e Sánchez-Crespo, 1994, páx. 20; Butcher, 1973, páx. 28; Downie e Heath, 1983, páx. 171; León e Montero, 1997, páxs. 80-81). Foron 17 os centros cos que foi posible traballar (Táboa 5.17.), dos cales 9 eran de ensinanza primaria e 6 eran de ensino secundario exclusivamente, mentres que outros dous eran tanto de ensinanza primaria coma de secundaria (ámbolos dous de titularidade privada). O alumnado procedía de 7 centros rurais, todos eles públicos (3 da costa e 4 de interior), ademais de 10 centros urbanos, dos que 4 eran da costa (3 públicos e 1 privado) e 7 de interior (5 públicos e 1 privado).

Táboa 5.17.: Número de centros que compoñen a mostra total

	Rural		Urbano		Total
	Pública	Privada	Pública	Privada	
Costa	3	---	3	1	7
Interior	4	---	5	1	10
Total	7		10		17

Ante a imposibilidade práctica de que o autor puidese efectuar a administración das probas persoalmente, pedíuselle ós orientadores e orientadoras dos centros que levaran a cabo esta tarefa. Coa finalidade de que dispuxeran do coñecemento necesario para isto, así como de establecer uns criterios que uniformaran o proceso, proporcionáronselle instrucións mediante entrevistas individuais. De tódolos modos, é preciso ter presente que utilizando este procedemento de actuación, non é posible garantir na súa totalidade a homoxeneidade e uniformidade do proceso e das condicións de aplicación. Este feito puido influír, provocando que as porcentaxes de protocolos válidos non fosen equivalentes nos diferentes centros e grupos.

A aplicación foi levada a cabo entre outubro de 2001 e febreiro de 2002, agás en dous centros por diferentes motivos, un de ensinanza primaria e outro de ensino secundario, nos que se desenvolveu no mes de abril e maio de 2002, respectivamente. Foi realizada en colaboración co profesorado que rexentaba a titoría, salvo nun de ensino secundario que administraron as probas a profesora de ciencias sociais e a de lingua galega. Aplicouse a Batería 1 e, transcorrida unha semana, a Batería 2.

Administráronselle as dúas Baterías a un total de 1116 alumnos e alumnas. Destes 1116 escolares, 599 cursaban sexto de Educación Primaria e 517 segundo de Educación Secundaria Obrigatoria; 373 pertencían ó ámbito rural e 743 ó urbano; 468 a zona de costa

e 648 de interior; 871 a centros públicos e 245 a centros de titularidade privada (Táboa 5.18.).

Táboa 5.18.: Número de escolares da mostra total inicial

		6.º Educación Primaria	2.º ESO	Totais
Hábitat 1	Rural	203	170	373
	Urbano	396	347	743
Hábitat 2	Costa	268	200	468
	Interior	331	317	648
Titularidade do centro	Público	479	392	871
	Privado	120	125	245
Totais		599	517	1116

Do mesmo xeito que na aplicación piloto e como xa se explicou no punto correspondente, excluíuse ó alumnado que repetía ou repetira algún curso (54 en sexto de Educación Primaria, o 9,01%, e 106 en segundo de ESO, o 20,50%), ó que presentaba dificultades importantes de lingua escrita á xuízo do profesorado (32 en sexto de Educación Primaria e 19 en segundo de ESO, que representan o 5,34% e o 3,67%, respectivamente), así como, nun segundo momento, ó alumnado que non puntuaba porque sinalaba un excesivo número de pseudopalabras. Respecto a esto último, como non era posible unha segunda administración, excluíronse 97 en Educación Primaria e 46 en ESO, que se corresponden respectivamente co 16,19% e co 8,90% de cada mostra.

Finalmente, excluíuse ó alumnado para o que o resultado da estimación nunha das Baterías resultaba ser o dobre ou máis que na outra (103 en Educación Primaria e 84 en

Educación Secundaria Obrigatoria; o 17,19% e o 16,25%, respectivamente). Excluíronse tendo en conta que a proporción dos que a puntuación era o dobre ou máis na Batería 1 respecto da segunda foi semellante á dos que ocorría á inversa e ante a posibilidade de que estes escolares actuaran en cada Batería dun modo moi diferente en canto á actitude e grao de motivación á hora de afrontar a tarefa.

Efectuadas estas exclusións, a porcentaxe media de suxeitos aceptado en 6.º de Educación Primaria foi do 53,90%, mentres que en ESO foi do 52,37% (Táboas 5.19. e 5.20.). Os excluídos porque sinalaron un excesivo número de pseudopalabras e, en consecuencia, non conseguían unha puntuación de 1 cando menos, ou porque acadaban unha puntuación o dobre ou máis nunha das Baterías que na outra foi do 33,38% en Educación Primaria e do 25,14% en ESO. O resto dos que se excluíron correspóndense con escolares repetidores ou con dificultades importantes en lectura (14,35% en Educación Primaria e 24,27% en ESO).

A mostra total resultante ou definitiva coa que se fixeron tódalas análises, unha vez efectuados os descartes anteriormente mencionados, quedou constituída por 575 escolares, 313 de sexto de Educación Primaria e 262 de segundo de Educación Secundaria Obrigatoria; 205 do ámbito rural e 370 do urbano; 262 de zona costeira e 313 de interior; 483 de públicos e 92 de centros privados (Táboa 5.21.). Do alumnado de sexto de Educación Primaria, 116 pertencían á zona rural e 197 á urbana; 156 á zona costeira e 157 á de interior; 271 a escolas públicas e 42 a centros de titularidade privada; 76 tiñan o galego como lingua familiar habitual, 116 o castelán e 121 as dúas. O alumnado de 2.º de ESO quedou distribuído do seguinte modo: 89 do rural e 173 de medio urbano; 106 de centros de costa e 156 de centros de interior; 212 de centros públicos e 50 de centros privados; 79 usaban o galego como lingua habitual na familia, 124 o castelán e 59 as dúas linguas.

Táboa 5.19. : Escolares excluídos e incluídos na mostra final de 6.º de Educación Primaria nos distintos centros

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	Total
***	7	3	0	0	0	2	0	0	1	0	0	
**	5	1	0	13	4	2	0	3	6	0	7	
*	5	7	4	4	4	0	0	2	1	5	0	
<	1	5	4	9	2	17	15	3	0	38	3	
>2	6	11	12	12	9	18	13	3	4	7	8	
Válidos	25	20	27	31	48	38	29	26	24	13	32	313
%Válidos	51,02	42,55	57,44	44,93	68,57	49,35	48,33	70,27	72,73	19,70	68,08	53,90

*** repetidores con dificultades en lectura ** repetidores sen dificultades * alumnado sen repetir, pero con dificultades en lectura < puntúan “cero” pero non teñen dificultades <2 diferenza do dobre ou máis nos resultados da estimación do nº de palabras coñecidas entre a 1ª e a 2ª batería *Válidos* casos aceptados despois de eliminar a tódolos que presentan algunha das características anteriores

Táboa 5.20.: Escolares excluídos e incluídos na mostra final de 2.º de ESO nos distintos centros

	C1	C2	C3	C4	C5	C6	C7	C8	Total
***	10	0	2	0	0	7	0	5	
**	12	16	10	22	5	3	5	9	
*	0	0	11	0	0	8	0	0	
<	4	2	5	7	9	0	19	0	
<2	13	4	11	9	21	3	16	7	
Válidos	32	38	30	55	25	31	25	26	262
% válidos	45,07	70,37	40	67,9	35,21	63,26	33,78	63,41	52,37

*** repetidores con dificultades en lectura ** repetidores sen dificultades *alumnado sen repetir, pero con dificultades en lectura < puntúan “cero” pero non teñen dificultades <2 diferenza do dobre ou máis nos resultados da estimación do nº de palabras coñecidas entre a 1ª e a 2ª batería *Válidos* casos aceptados despois de eliminar a tódolos que presentan algunha das características anteriores

Táboa 5.21.: Número de escolares que compoñen a mostra total definitiva

		6.º Educación Primaria	2.º ESO	Totais
Hábitat 1	Rural	116	89	205
	Urbano	197	173	370
	Total	313	262	575
Hábitat 2	Costa	156	106	262
	Interior	157	156	313
	Total	313	262	575
Titularidade do centro	Público	271	212	483
	Privado	42	50	92
	Total	313	262	575
Lingua habitual	Galego	76	79	155
	Castelán	116	124	240
	As dúas	121	59	180
	Total	313	262	575

O alumnado de 6.º de Educación Primaria procedía de 11 centros (Táboa 5.22.). Catro eran rurais públicos (2 de costa e 2 de interior) e 7 urbanos : 5 públicos (2 de costa e 3 de interior) e 2 privados (1 de costa e 1 de interior). En canto á mostra de 2.º de ESO (Táboa 5.23.), os centros de procedencia por categorías foron: 3 rurais (1 de costa e 2 de interior) e 5 urbanos (3 públicos, sendo 1 de costa e 2 de interior, e 2 privados, emprazado 1 en zona de costa e 1 en zona de interior).

Táboa 5.22.: Centros de procedencia do alumnado de 6.º de Educación Primaria segundo o hábitat e a titularidade do centro

	Rural		Urbano		Total
	Pública	Privada	Pública	Privada	
Costa	2	----	2	1	5
Interior	2	----	3	1	6
Total	4		7		11

Táboa 5.23.: Centros de procedencia do alumnado de 2.º de ESO segundo o hábitat e a titularidade do centro

	Rural		Urbano		Total
	Pública	Privada	Pública	Privada	
Costa	1	----	1	1	3
Interior	2	----	2	1	5
Total	3		5		8

Non se deben esquecer as limitacións que se foron expoñendo no relativo á mostraxe en relación cos descartes de suxeitos que se efectuaron, a que non se conseguira unha mostra aleatoria nin proporcional ou con que o número de suxeitos por agrupamento ou cela non fose suficiente. Por iso, farase necesario interpretar con cautela os resultados das análises que se realicen. Como en todo traballo experimental, será preciso seguir investigando para tratar de solucionar este tipo de limitacións, pero tendo presente que o test obxecto deste traballo conta xa cun considerable apoio histórico.

5.4 Fiabilidade

Se unha persoa se pesa nunha báscula repetidamente (pésase, báixase da báscula, vólvese a pesar ó cabo duns momentos ..., sen move-la báscula, sen cambia-la cantidade de roupa, etc.) e obtén as mesmas cantidades de peso, conclúese que o instrumento é fiable, confiable, preciso. Os tests psicolóxicos e pedagóxicos teñen que ser fiables, aínda que non sempre sexan tan confiables coma unha báscula coma a do exemplo; isto é, están sometidos a algún erro de medida.

Na literatura de construción de tests (Fernández Ballesteros e Carroble, 1983; Kline, 1986; Muñiz, 2000) distínguense catro métodos para calcula-la fiabilidade dun test: consistencia interna, dúas metades, test-retest e formas paralelas. O método test-retest aborda máis intuitivamente a estabilidade das puntuacións: correspóndese co exemplo co que comezou este apartado. O das dúas metades pódese incluír dentro do de consistencia interna

Para calcula-la consistencia interna dun test utilízase habitualmente o *coeficiente alfa* de Cronbach, que Kline (1986, p. 15) define como “a correlación estimada do test con outro test da mesma lonxitude procedente do mesmo universo de ítems”. A estimación aludida realízase a partir do grao en que varían e covarían os ítems que constitúen o test, polo que se basea nunha soa aplicación do instrumento.

A fiabilidade en termos de estabilidade ou constancia das puntuacións comunmente calcúlase, ben repetindo a aplicación do mesmo test (test-retest), o que presenta inconvenientes (por exemplo, uns suxeitos pódense beneficiar máis que outros da primeira aplicación na segunda), ben administrando formas paralelas do mesmo test, o cal tamén ten as súas dificultades, principalmente porque non é fácil construír formas equivalentes. Se se opta polo test-retest, tómanse medicións dos suxeitos en momentos distintos, o que tamén pode ocorrer coa aplicación de formas paralelas.

As estimacións da fiabilidade baséanse na teoría do erro de medida. Dous feitos importantes que hai que ter en conta é que o coeficiente de fiabilidade aumenta cando aumenta a variabilidade da mostra de suxeitos cos que se calculou e tamén co incremento do número de ítems do test (Muñiz, 2000, pp. 45 e 47).

Os tests deben posuír unha fiabilidade aceptable para que sexan útiles, tanto se a súa finalidade é a investigación coma se é a aplicación práctica. Un coeficiente considérase aceptable cando se sitúa entre 0,80 e 1 (Anastasi, 1977, p. 80).

Neste traballo, para estima-la fiabilidade, primeiramente, optouse polo procedemento de correlación entre dúas formas paralelas, utilizando o *coeficiente de Pearson*, en tanto que, por un lado, contábase con dúas Baterías que se supoñían paralelas con datos relativos á estimación do vocabulario coñecido, de modo que se podían utiliza-la correlacións entre elas como unha estimación deste tipo de fiabilidade.

No que se refire á *mostra total* (N=575), utilizando a estimación do número de palabras coñecidas mediante as Baterías, a fiabilidade parece aceptable (Táboa 5.24., p. 245), xa que a correlación é suficientemente alta entre Baterías 1 e 2 (0,806).

Non obstante, unha comparación mediante a proba *t-test* das medias obtidas con ámbalas dúas Baterías indica que son significativamente diferentes ($t_{(574)} = -7,181$; $p = 0,000 < 0,05$). Por este motivo, as estimacións do vocabulario total resultan un pouco distintas (superiores as baseadas na Batería 2), aínda que a orde relativa dos suxeitos tenda a se-la mesma nas dúas. Non sabemos ata que punto puido influír que a Batería 2 sempre se aplicara en segundo lugar, non de forma contraequilibrada. Como consecuencia do dito, en tanto non se investigue mellor esta cuestión, débese indicar en cal das dúas Baterías está baseada a estimación.

Outra forma de aborda-la fiabilidade é considera-las dúas Baterías como dúas metades dun único test. Aplicándolle ós datos á *fórmula de Spearman-Brown* ($R = 2r_{xx}/(1 + r_{xx})$), para un test

de lonxitude dobre ($2 \times 0,86 / 1 + 0,86$), obtense unha fiabilidade de 0,892, máis alta que a acadada considerando as dúas Baterías como formas paralelas, como cabe esperar do principio xeral de que a fiabilidade aumenta cando se aumenta a lonxitude do test. Parece que resulta máis fiable o test se se efectúa a estimación do vocabulario en base á aplicación das dúas Baterías.

Como información complementaria preséntanse as intercorrelacións entre os subtests das dúas Baterías. Estas correlacións son significativas na súa totalidade e, como se pode observar na táboa 5.24., sitúanse na súa meirande parte entre 0,25 e 0,50, se ben, hai tres que son inferiores a 0,25 (entre o 1.º subtest e o 5.º subtest, entre o 2.º e o 4.º e entre o 2.º e o 5.º) e unha que supera 0,50 (entre o 8.º subtest e o 9.º, que é de 0,531).

Polo que respecta á mostra de *6.º de Educación Primaria* ($N=313$), as correlacións das puntuacións correspondentes á estimación do número de palabras coñecidas entre as Baterías é de 0,793 (Táboa 5.25., p. 246). Tamén neste caso son significativamente diferentes as estimacións das dúas Baterías ($t_{(312)} = -4,590$; $p = 0,000 < 0,05$). Polo tanto, caben as mesmas consideracións que as feitas para a mostra total. Se se calcula a fiabilidade entendendo as dúas Baterías como dúas metades dun só test mediante a fórmula de Spearman-Brown, o coeficiente sería de 0,88.

En canto á información complementaria sobre as correlacións entre os diferentes subtests, estas están comprendidas entre 0,20 e 0,50, segundo se pode comprobar na táboa citada con anterioridade, agás catro que se sitúan por debaixo de 0,20 (1.º co 4.º subtest e 1.º co 5.º subtest, 2.º subtest co 5.º e 5.º subtest co 10.º) e unha que é superior a 0,50 (8.º co 9.º, que resulta ser 0,531), dándose o caso, ademais, de que a única que non é significativa é a que lle corresponde ó 1.º subtest co 5.º.

Finalmente, no relativo á *mostra de 2.º de Educación Secundaria Obrigatoria* (N=262), a correlación entre a estimación do número de palabras coñecidas nas dúas Baterías é de 0,807 (Táboa 5.26., p. 247). Nesta mostra son igualmente significativas as diferenzas entre as estimacións das dúas Baterías ($t_{(261)} = -5,572$; $p = 0,000 < 0,05$). Por esta razón, son tamén válidas as consideracións realizadas para a mostra total. A fiabilidade calculada coa fórmula de Spearman-Brown foi de 0,893.

Na Táboa 5.26. pódese comprobar que as correlacións entre os diferentes subtests son todas significativas. Sitúanse fundamentalmente entre 0,20 e 0,50, coas excepcións das que corresponden entre o 1.º e o 5.º, que é inferior a 0,20 e as que corresponden ó 3.º co 4.º, o 3.º co 9.º e o 4.º co 9.º, que superan 0,50 (0,542, 0,52 e 0,531, respectivamente).

Táboa 5.24.: Correlacións utilizando a estimación do número de palabras coñecidas para a mostra total (N=575)

correlación significativa (bilateral)		Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Batería 1	Batería 2	** A é ó 0,01%
	Test 1	0,274**	0,275**	0,248**	0,145**	0,302**	0,314**	0,337**	0,337**	0,270**		0,445**	
	Test 2		0,279**	0,224**	0,201**	0,287**	0,348**	0,355**	0,418**	0,243**		0,474**	
	Test 3			0,474**	0,329**	0,360**	0,464**	0,460**	0,506**	0,371**		0,616**	
	Test 4				0,266**	0,331**	0,403**	0,496**	0,486**	0,401**		0,601**	
	Test 5					0,285**	0,348**	0,343**	0,367**	0,251**		0,451**	
	Test 6						0,326**	0,326**	0,377**	0,254**	0,494**		
	Test 7							0,424**	0,462**	0,328**	0,591**		
	Test 8								0,531**	0,394**	0,625**		
	Test 9									0,416**	0,660**		
	Test 10										0,490**		
	Batería 1											0,806**	

Táboa 5.25.: Correlacións utilizando a estimación do número de palabras coñecidas para a mostra de 6.º de Educación Primaria (N=313)

	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Batería 1	Batería 2
Test 1	0,266**	0,264**	0,189**	0,095	0,243**	0,240**	0,257**	0,311**	0,248**		0,394**
Test 2		0,276**	0,199**	0,180**	0,326**	0,256**	0,344**	0,418**	0,224**		0,473**
Test 3			0,397**	0,322**	0,372**	0,465**	0,420**	0,476**	0,308**		0,608**
Test 4				0,237**	0,306**	0,328**	0,498**	0,433**	0,417**		0,583**
Test 5					0,275**	0,337**	0,320**	0,303**	0,192**		0,420**
Test 6						0,292**	0,303**	0,365**	0,213**	0,502**	
Test 7							0,329**	0,432**	0,276**	0,537**	
Test 8								0,532**	0,343**	0,599**	
Test 9									0,411**	0,637**	
Test 10										0,454**	
Batería 1											0,793**

** A correlación é significativa ó 0,01% (bilateral)

Táboa 5.26.: Correlacións utilizando a estimación do número de palabras coñecidas para a mostra de 2.º de ESO (N=262)

	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10	Batería 1	Batería 2
Test 1	0,249**	0,263**	0,292**	0,188**	0,361**	0,356**	0,396**	0,336**	0,276**		0,473**
Test 2		0,249**	0,230**	0,217**	0,244**	0,409**	0,328**	0,374**	0,239**		0,442**
Test 3			0,542**	0,330**	0,346**	0,444**	0,483**	0,520**	0,424**		0,612**
Test 4				0,290**	0,353**	0,456**	0,485**	0,531**	0,376**		0,611**
Test 5					0,293**	0,354**	0,362**	0,434**	0,303**		0,479**
Test 6						0,358**	0,351**	0,395**	0,292**	0,491**	
Test 7							0,485**	0,461**	0,355**	0,616**	
Test 8								0,500**	0,430**	0,633**	
Test 9									0,404**	0,667**	
Test 10										0,512**	
Batería 1											0,812**

** A correlación é significativa ó 0,01% (bilateral)

En términos xerais e coa precaución que aconsellan as limitacións da mostraxe, en base ás correlacións obtidas mediante a fórmula de Pearson, parece razoable concluír que os coeficientes son aceptables, xa que son significativas ó n. de s. do 1% e están en torno ó 0,80. Pero, debe terse en conta ó respecto que as estimacións efectuadas mediante as dúas Baterías son significativamente diferentes. Agora ben, se se calcula a fiabilidade mediante a fórmula de Spearman-Brown, considerando as dúas Baterías como dúas metades dun test, as correlacións elévanse ata arredor de 0,89.

5.5 Validez

5.5.1 Consideracións sobre a validez

Podemos defini-la validez como o grao en que o instrumento mide o que se supón que debe medir (Seoane e Rechea, 1976, p. 59), é dicir, se é útil para a finalidade que se elaborou.

Existen diferentes tipos de validez que implican distintos procedementos para determinala, se ben todos se basean nas relacións entre a actuación nun test e outros feitos observables relativos á conduta que se está considerando. Tamén varía duns autores a outros, tanto a denominación dalgúns dos tipos coma o número dos mesmos que establecen (Anastasi, 1977; Cabrera e Espín, 1986; Cohen, Manion e Morrison, 2001; Fernández Ballesteros e Carroble, 1983; Muñiz, 2000; Rust e Golombok, 1995; Seoane e Rechea, 1976). Con todo, hai tres que se contemplan habitualmente en tódolos casos:

1. Validez de construto. O propósito é valida-la teoría en que se basea a proba elaborada e refírese ó grao en que mide unha variable ou trazo definidos previamente. Sería a principal forma de validación (Rust e Golombok, 1995, p. 80) e abarca o estudo de calquera outra clase de validez (Seoane e Rechea, 1976, p. 62).

Un dos procedementos básicos consiste en:

- Identificar teoricamente os factores que implica a execución do test, a teoría de que parte.
- Establecer hipóteses relativas ó trazo que se mide.
- Verifica-las hipóteses.

Por elo, este tipo de validez vaise elaborando a medida que se estuda e investiga o test e nunca se completa, converténdose nun continuo de recollida de información co obxecto de esclarece-la natureza do construto que consideramos (Cabrera e Espín, 1986; Rust e Golombok, 1995).

A información que debe proporcionar o manual dun test para xustificar esta validez será, segundo Cabrera e Espín (1986, p. 126):

- Datos relativos á teoría psicolóxica que fundamenta o test.
- Datos referentes ás características dos suxeitos que poden afectar ó trazo avaliado (idade, sexa, nivel socioeconómico ...).
- Información bibliográfica e investigacións realizadas co test, que permitan esclarece-la natureza do atributo que se mide.

2. Validez de contido. Supón o exame sistemático do contido do test para determinar se comprende unha mostra representativa do contido que pretende medir ou avaliar, xa que fai referencia á garantía desta adecuación e representatividade.

Esixe unha análise sistemática da área a avaliar, de forma que tódolos aspectos estean cubertos e nas proporcións correctas (Anastasi, 1977; Cabrera e Espín, 1986; Cohen, Manion e Morrison, 2001; Fernández Ballesteros e Carrobbles, 1983; Muñiz, 2000). Para iso, debería utilizarse unha mostraxe aleatoria dos elementos dun dominio definido, o que na práctica non sempre resulta factible. Se se logra e dispoñemos de dúas formas paralelas do instrumento construídas independentemente, por medio da mostraxe de elementos, pódese

calcula-lo coeficiente de validez comparando as puntuacións dunha mostra de suxeitos nas dúas formas do test (Fernández Ballesteros e Carrobles, 1983). Hai dous aspectos importantes a conseguir (Livingston, 1972):

- Que o instrumento non inclúa aspectos irrelevantes do contido.
- Que o instrumento inclúa tódolos aspectos importantes do dominio.

Dada a carencia de índices e instrumentos para determina-la validez de contido, debe realizarse por medio de procedementos racionais máis que empíricos, que pasan por análises dos contidos posibles e consultas a expertos na área que se aborda. A isto é preciso engadirlle a necesidade de tratar de que todo tipo de contidos estean presentes na adecuada proporción de ítems (Muñiz, 2000). Tamén se pode completar con procedementos empíricos, como é verifica-lo progreso no grao de dominio dos elementos do test, é dicir, se co aumento de nivel de escolarización se conseguen mellores resultados (Anastasi, 1977). Normalmente selecciónanse os elementos nos que se produce unha maior diferenza nas porcentaxes de suxeitos que os superan desde os niveis escolares máis baixos ós superiores.

É importante ser prudentes na tendencia a xeneraliza-lo resultado obtido. No tema que nos ocupa, o rendemento en algo específico, como a estimación do número de palabras coñecidas mediante unha proba, non permite deducir que ese vocabulario vaia a ser "activo". Alén disto, o ser unha proba escrita, este é un factor que pode influír no rendemento, xa que esixe exactitude, velocidade e comprensión lectoras, cando menos de palabras.

Relacionada con este tipo de validez está a *Validez Aparente*, que a veces se inclúe como un tipo especial da de contido (Muñiz, 2000). Non se refire ó que realmente mide o test, senón ó que parece medir, a que lle pareza a quen se lle aplica que é adecuado para medi-lo que se pretende. Pode ter importancia cara a motivación e actitude dos suxeitos, xa que senón lle parecen adecuados os ítems, probablemente non se implicarán na súa correcta realización co gallo de obte-la mellor puntuación posible.

O estudo sobre este tipo de validez debería incluír información sobre as áreas, habilidades ou os obxectivos que abarca o test, con indicación dos elementos de cada categoría. Así mesmo, debería describi-lo procedemento para a selección das categorías e clasificación dos elementos e, se participaron xuíces nela, indicalo, así como as instrucións que recibiron para clasifica-los citados elementos e o grao de acordo. Tamén é importante proporcionar información sobre as fontes de onde se extraeron os contidos.

3. Validez empírica ou relacionada cun criterio. Indica a eficacia do test na predición de determinada variable, a partir do rendemento no test. Para calculala, compárase este rendemento cun criterio ou medida distinta do aspecto que se pretende avaliar co test (Anastasi, 1977; Cohen, Manion e Morrison, 2001; Fernández Ballesteros e Carrobles, 1983; Muñiz, 2000; Seoane e Rechea, 1976). Ponse o énfase en que medida a proba serve para predicir unha conduta futura, ou como medida substitutiva dunha execución ou desempeño externo (Cabrera e Espín, 1986). É de utilidade cando interesa coñecer se o instrumento discrimina entre grupos de suxeitos clasificados de acordo a criterios independentes.

Os estatísticos utilizados para estimala son os mesmos que para estima-la fiabilidade. Segundo o deseño que se utilice para calcula-lo coeficiente de validez, pódese diferenciar entre:

- a) Validez concorrente, cando o test e o criterio se miden ó mesmo tempo. Fai referencia á concordancia entre as probas que pretenden medi-lo mesmo aspecto.
- b) Preditiva, cando o criterio se mide nun período de tempo posterior (no caso de face-lo antes, sería retrospectiva). Refírese ó grao en que coa medida obtida se pode estima-lo rendemento futuro dun suxeito.

O problema fundamental é establecer e medi-lo criterio. Entre os criterios máis utilizados están o rendemento académico, outro test coñecido que mida o mesmo contido ou o emprego de grupos contrastados.

Se un índice de validez basicamente representa o grao en que unha proba mide o que di medir, é necesario ser conscientes de que no relativo ó ámbito de coñecemento que se aborda, así como á intencionalidade do test construído (estimación do vocabulario coñecido), polo momento non se conseguiu establecer un criterio claramente obxectivo que se relacione claramente con este construto. Aínda tendo presente esta dificultade, intentouse valida-lo test, en base ás consideracións e datos que a continuación se comentan.

5.5.2 Validez de construto

Polo que se refire á **validez de construto**, neste traballo, en primeiro lugar, en termos xerais apóiase na achega teórica exposta ó longo da primeira parte, así como no apartado relativo á descrición da mostra desta segunda parte (apartado 5.3).

En segundo lugar, en relación co concepto de vocabulario, tendo en conta a propia natureza do construto, así como toda a información que se achegou ó respecto nos capítulos 1 e 2 e no apartado 5.2 sobre a situación lingüística en Galicia, cabe a posibilidade de formular diferentes hipóteses no relativo ó seu tamaño. Por exemplo:

- a) Que aumente coa idade, como reflexo do progresivo incremento na adquisición de coñecementos do mundo.
- b) Que exista unha relación co rendemento escolar en xeral, pero sobre todo con determinadas áreas, como é a comprensión lectora.
- c) Que exista relación coa intelixencia, en tanto que é unha proba que habitualmente se inclúe en tests de intelixencia.
- d) Que correlacione positivamente co uso dunha lingua.
- e) Que exista unha correlación positiva coa fluencia verbal ou coa riqueza de asociacións verbais.
- f) Que correlacione diferencialmente consonte ós niveis académicos dos proxenitores do alumnado.

A información acadada acerca da utilización e competencia en lingua galega por parte da poboación galega, exposta no apartado 5.2. deste capítulo, non sempre parece ir no mesmo sentido, en relación coas dicotomías establecidas para os estratos rural/urbano e a clase social, xa que, como vimos, en xeral a poboación rural utilízaa máis, pero síntese menos competente que a urbana, do mesmo modo que ocorre entre a clases baixa/media-baixa e a media-alta.

Aínda tendo en conta isto, e que poderíamos pensar que unha maior competencia na escrita implicaría un maior tamaño do vocabulario, en tanto que a lectura é unha fonte esencial no seu incremento, tamén é certo que se esta maior competencia non se utiliza, é dicir, non se le realmente, dificilmente repercutirá neste sentido.

Agora ben, a teor dos devanditos datos, co obxecto de estudar a validez do instrumento elaborado, formuláronse cinco hipóteses no sentido de que o test elaborado será sensible ás diferenzas no tamaño do vocabulario entre o alumnado dos diferentes grupos establecidos.

As cinco hipóteses direccionais contempladas son as seguintes:

$$1.^a \quad H_0: \mu_{2.^\circ ESO} - \mu_{6.^\circ E. Primaria} \leq 0$$

$$H_1: \mu_{2.^\circ ESO} - \mu_{6.^\circ E. Primaria} > 0$$

$$2.^a \quad H_0: \mu_{Rural} - \mu_{Urbano} \leq 0$$

$$H_1: \mu_{Rural} - \mu_{Urbano} > 0$$

$$3.^a \quad H_0: \mu_{Interior} - \mu_{Costa} \leq 0$$

$$H_1: \mu_{Interior} - \mu_{Costa} > 0$$

$$4.^a \quad H_0: \mu_{Baixa e Medio-baixo} - \mu_{Medio-alto} \leq 0$$

$$H_1: \mu_{Baixa e Medio-baixo} - \mu_{Medio-alto} > 0$$

$$5.^a \quad H_0: \mu_{Lingua galega} - \mu_{Lingua castelán/As dúas} \leq 0$$

$$H_1: \mu_{Lingua galega} - \mu_{Lingua castelán/As dúas} > 0$$

Primeira hipótese, que será a hipótese básica, refírese ós agrupamentos por idade e nivel de escolarización: O tamaño estimado do vocabulario coñecido polo alumnado de 2.º de ESO é maior que o tamaño estimado do vocabulario coñecido polo alumnado de 6.º de Educación Primaria.

Segunda hipótese, referida ás diferenzas entre hábitat rural e urbano: Os escolares do ámbito rural obterán puntuacións máis altas na estimación do vocabulario que os de ámbito urbano.

Terceira hipótese, relativa ás diferenzas entre hábitat de interior e costeiro: Será maior o vocabulario estimado do alumnado que pertence a unha zona de interior que o do alumnado que pertence a unha zona de costa.

Cuarta hipótese, establece un prognóstico en relación co nivel socioeconómico, tendo en conta as matizacións coas que se concibiu a mostra ó respecto, é dicir, a equivalencia dos niveis en base á titularidade do centro: O alumnado de nivel socioeconómico baixo/medio-baixo (centros públicos) obterá unha estimación do vocabulario maior que o alumnado de nivel medio-alto (centros privados).

Quinta hipótese, establece unha predición relacionada coa lingua familiar usual: O vocabulario do alumnado que ten o galego como lingua habitual ou preferente no medio familiar será maior que o do alumnado que ten o castelán como lingua familiar e o do alumnado que ten as dúas.

Para a comprobación da existencia de diferenzas, nun primeiro momento efectuáronse unha serie de *ANOVAS* de 2 factores, utilizando como factores fixos, por unha parte, o nivel de escolarización/idade e, por outra, ben o hábitat rural/urbano (Táboas 5.35., 5.36. e 5.37.), ben o hábitat costa/interior (Táboas 5.38., 5.39. e 5.40.), ben o nivel

socioeconómico (Táboas 5.41., 5.42. e 5.43.) ou ben a lingua usual na familia (Táboas 5.44., a 5.49.).

Para poder utiliza-la técnica *ANOVA* (Downie e Heath, 1983; Tejedor, 1984, entre outros), deben cumprirse tres condicións:

- a) Que a distribución das mostras se corresponda coa distribución normal.
- b) Que a varianza das mostras sexa homoxénea.
- c) Que as mostras sexan independentes.

No relativo á distribución expóñense os resultados da análise de bondade de axuste á distribución normal dos datos referidos ó total da mostra, así como á de 6.º de Educación Primaria e de 2.º de ESO. Esta análise efectuouse mediante a Proba de Kolgomorov-Smirnov, dado que tódalas frecuencias esperadas resultaron ser inferiores a 5 (Downie e Heath, 1983; Seoane e Rechea, 1976; Tejedor, 1984). Con todo, para facer máis fluída a lectura do texto, aquí soamente se ofrecen de forma gráfica datos relacionados coa mostra total.

Polo que se refire á *mostra total*, os datos de significación non permiten rexeita-la hipótese nula, referida a que as puntuacións proceden dunha poboación normal. Pódese afirmar, polo tanto, que a distribución das puntuacións axústase á distribución normal. Só se presentan de forma máis completa nas Táboas 5.27. e 5.28. e na Figura 5.12. os datos relativos ó Total do Test, co obxecto sinalado anteriormente de facilita-la lectura do texto (Batería 1: Sig. asintótica= 0,439>0,05; Batería 2: Sig. asintótica= 0,412>0,05; Total Test: Sig. asintótica= 0,471>0,05).

Táboa 5.27.: Estatísticos descritivos da estimación do vocabulario para a mostra total

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	575	206	12991	5721,40	2539,64	105,91
Batería 2	575	206	14538	6226,35	2808,31	117,11
Test	575	206	13507	5973,63	2541,20	105,97
N válido (segundo lista)	575					

Táboa 5.28.: Proba de Kolgomorov-Smirnov para o Total do Test na mostra total

		Total Test
N		575
Parámetros normais a,b	Media	5973,63
	Desviación típica	2541,197
Diferenzas máis extremas	Absoluta	,035
	Positiva	,029
	Negativa	-,035
Z de Kolgomorov-Smirnov		,847
Sig. Asintót. (bilateral)		,471

a. A distribución de contraste é a Normal.

b. Calculouse a partir dos datos.

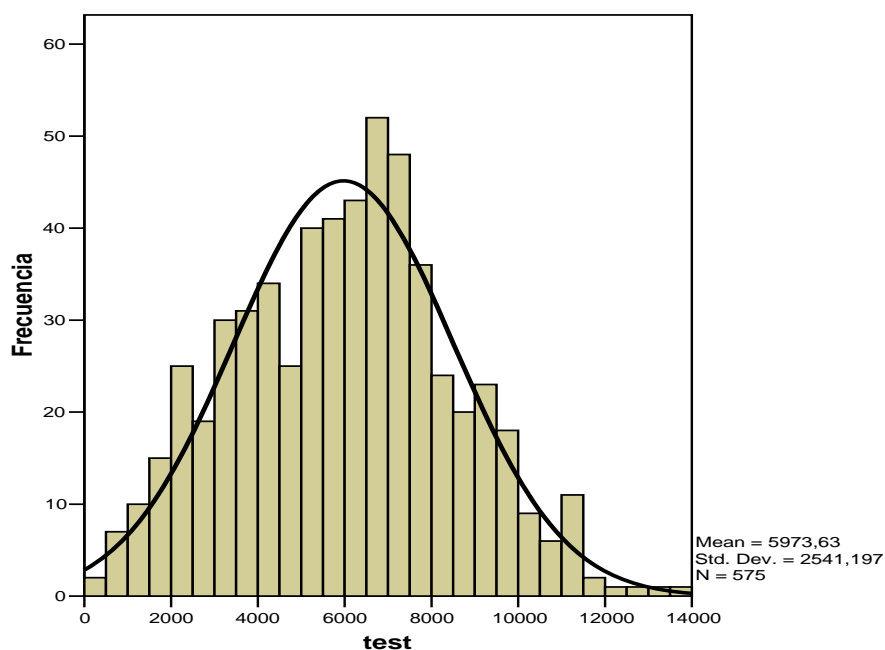


Figura 5.12.: Distribución da estimación do vocabulario para o Total do Test na mostra total

En canto á *mostra de 6.º de Educación Primaria*, unicamente se expoñen os estatísticos descritivos na Táboa 5.29. Tampouco neste caso os datos de significación permiten rexeita-la hipótese nula respecto da normalidade da distribución (Batería 1: Sig. asintótica= 0,571>0,05; Batería 2: Sig. asintótica= 0,527>0,05; Total Test: Sig. asintótica= 0,371>0,05).

Táboa 5.29.: Estatísticos descritivos da estimación do vocabulario na mostra de 6.º de Educación Primaria

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	313	206	10826	5342,17	2380,21	134,54
Batería 2	313	206	12373	5761,20	2595,69	146,72
Test	313	206	11187	5551,35	2355,79	133,16
N válido (segundo lista)	313					

Polo que respecta á *mostra de 2.º de ESO*, pódese afirmar, como nos casos anteriores, que a distribución das puntuacións axústase á distribución normal (Batería 1: Sig. asintótica= 0,576>0,05; Batería 2: Sig. asintótica= 0,847>0,05; Total Test: Sig. asintótica= 0,905>0,05). Tamén nesta ocasión se expoñen soamente os estatísticos descritivos (Táboa 5.30.).

Táboa 5.30.: Estatísticos descritivos da estimación do vocabulario na mostra de 2.º de ESO

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	262	618	12991	6174,45	2652,09	163,85
Batería 2	262	618	14538	6782,04	2953,12	182,44
Test	262	618	13507	6478,11	2664,28	164,60
N válido (segundo lista)	262					

Respecto da homoxeneidade de varianzas, os datos obtidos son dispares. Así, considerando o nivel de escolarización e o hábitat rural/urbano (Táboa 5.31.), para os datos da Batería 1 e do Total do Test si se pode asumir a esta homoxeneidade de varianzas, pero non é posible asumi-la para os datos da Batería 2 ($F_{(3)} = 2,981$; $p = 0,031 < 0,05$). Considerando o nivel de escolarización e o hábitat costa/interior (Táboa 5.32.), as diferenzas entre as varianzas son sempre significativas, tanto para os datos coa Batería 1 (Batería1: $F_{(3)} = 3,034$; $p = 0,029 < 0,05$), como coa Batería 2 ($F_{(3)} = 3,273$; $p = 0,021 < 0,05$) e co Total do Test ($F_{(3)} = 3,486$; $p = 0,016 < 0,05$). En canto á análise tendo en conta o nivel de escolarización e o nivel socioeconómico en base á titularidade dos centros (Táboa 5.33.), pódese manter que as varianzas son homoxéneas para os datos coa Batería 1 e co Total do Test, pero non coa Batería 2 ($F_{(3)} = 3,525$; $p = 0,015 < 0,05$). No contraste atendendo ó nivel de escolarización e á lingua familiar habitual (Táboa 5.34.), os datos mostran que as

varianzas son homoxéneas, tanto cos datos da Batería 1 como cos da Batería 2 e do Total do Test.

Táboa 5.31.: Contraste de Levene sobre homoxeneidade das varianzas (nivel de escolarización e hábitat rural vs urbano)

	F	gl1	gl2	Significación
Batería 1	2,241	3	571	0,083
Batería 2	2,981	3	571	0,031
Total Test	2,424	3	571	0,065

Táboa 5.32.: Contraste de Levene sobre homoxeneidade das varianzas (nivel de escolarización e hábitat costa vs interior)

	F	gl1	gl2	Significación
Batería 1	3,034	3	571	0,029
Batería 2	3,273	3	571	0,021
Total Test	3,486	3	571	0,016

Táboa 5.33.: Contraste de Levene sobre homoxeneidade das varianzas (nivel de escolarización e nivel socioeconómico-titularidade dos centros)

	F	gl1	gl2	Significación
Batería 1	0,854	3	571	0,465
Batería 2	3,525	3	571	0,015
Total Test	2,428	3	571	0,064

Táboa 5.34.: Contraste de Levene sobre homoxeneidade das varianzas
(nivel de escolarización e lingua familiar habitual)

	F	gl1	gl2	Significación
Batería 1	0,797	5	569	0,552
Batería 2	1,241	5	569	0,289
Total Test	2,428	5	569	0,576

É preciso sinalar que, se ben non se cumpría a condición de homoxeneidade de varianzas en tódolos casos, tendo en conta a opinión de autores como Ferguson (1971, p. 219) ou Lewis (1971, p. 8) utilizouse a técnica *ANOVA*. Isto afecta ós contrastes entre as mostras rural e urbana relativos á estimación do vocabulario coñecido mediante a Batería 2, entre as mostras costeira e de interior e entre centros públicos e privados no que se refire á estimación coa Batería 2.

Co fin de facilita-la lectura deste punto, os estatísticos básicos relativos á estimación do número de palabras coñecidas (medias, mínimos e máximos, desviacións típicas, etc.) presentaranse no capítulo 6 (Táboas 6.1. a 6.16.), adicado ós baremos, aínda que se fará referencia ás medias a medida que se vaia comentando o contraste das distintas hipóteses.

Os resultados da análise mediante *ANOVA* para a *mostra total*, como se pode comprobar nas Táboas 5.35. a 5.49., mostran que as diferenzas son significativas en tódolos casos, cando se consideran de forma independente os factores nivel de escolarización, hábitat rural/urbano, hábitat interior/costa, nivel socioeconómico e lingua familiar habitual. Neste último caso, unha comparación post hoc mediante a *proba de Scheffé*, permite concluír que é máis amplo o vocabulario dos escolares que teñen o galego como lingua habitual na familia (Táboas 5.45., 5.47. e 5.49.). En canto á interacción, os datos indican

que non é significativa entre o nivel de escolarización e o resto dos factores, aínda que en relación co hábitat costa/interior (Táboas 5.38., 5.39. e 5.40.; Figuras 5.16., 5.17. e 5.18.), non está lonxe de selo (Batería1: $F_{(1, 571)} = 3,108$; $p = 0,078$. Batería 2: $F_{(1, 571)} = 3,463$; $p = 0,063$. Total Test: $F_{(1, 571)} = 3,662$; $p = 0,056$). Pola súa parte, o tamaño do efecto é moi limitado en tódalas variables (sempre por debaixo do 5%).

No relativo á **primeira hipótese**, a media de palabras estimadas é maior no caso do alumnado de 2.º de ESO que a media dos alumnado de 6.º de Educación Primaria (Táboa 6.1.). Pódese entender que os datos que proporciona a *ANOVA* permiten mante-la hipótese e conseguintemente afirmar que o vocabulario do alumnado de 2.º de ESO é significativamente maior que o vocabulario do alumnado de 6.º de Educación Primaria, xa que as diferenzas de medias son estatisticamente significativas a un nivel de confianza do 5% (Táboas 5.35. a 5.49.).

As catro hipóteses restantes, refírense ós agrupamentos relacionados co medio socioxeográfico, co medio socioeconómico e coa lingua familiar habitual. Analizaranse as diferenzas para a mostra total. Pero tamén se comentarán a tendencia destas diferenzas para as mostras de 6.º de Educación Primaria e de 2.º de ESO por separado, en base ós gráficos que representan os datos da análise.

No que se refire á **segunda hipótese** (rural/urbano), en canto ó *total da mostra* ($N=575$), considerando os dous niveis escolares conxuntamente, son maiores as medias do vocabulario estimado dos escolares do ámbito rural (Táboa 6.2.). Pódese mante-la hipótese, é dicir, o tamaño do vocabulario do alumnado de zona rural é significativamente maior que o tamaño do vocabulario de alumnado de zona urbana, xa que as diferenzas entre as medias son significativas ó 5%, conforme se observa nas Táboas 5.35., 5.36. e 5.37. (Batería 1: $F_{(1, 571)} = 6,980$; $p = 0,008 < 0,05$. Batería 2: $F_{(1, 571)} = 6,952$; $p = 0,009 < 0,05$. Total Test: $F_{(1, 571)} = 7,731$; $p = 0,006 < 0,05$) .

Táboa 5.35.: Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Batería 1)

Pruebas de los efectos inter-sujetos

Variable dependiente: ba1

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	155547420 ^a	3	51849140	8,348	,000	,042
Intersección	1,781E+10	1	1,78E+10	2868,135	,000	,834
nivelesc	120167526	1	1,20E+08	19,347	,000	,033
habitat1	43356304,1	1	43356304	6,980	,008	,012
nivelesc * habitat1	18929616,8	1	18929617	3,048	,081	,005
Error	3546614302	571	6211233,5			
Total	2,252E+10	575				
Total corregida	3702161722	574				

a. R cuadrado = ,042 (R cuadrado corregida = ,037)

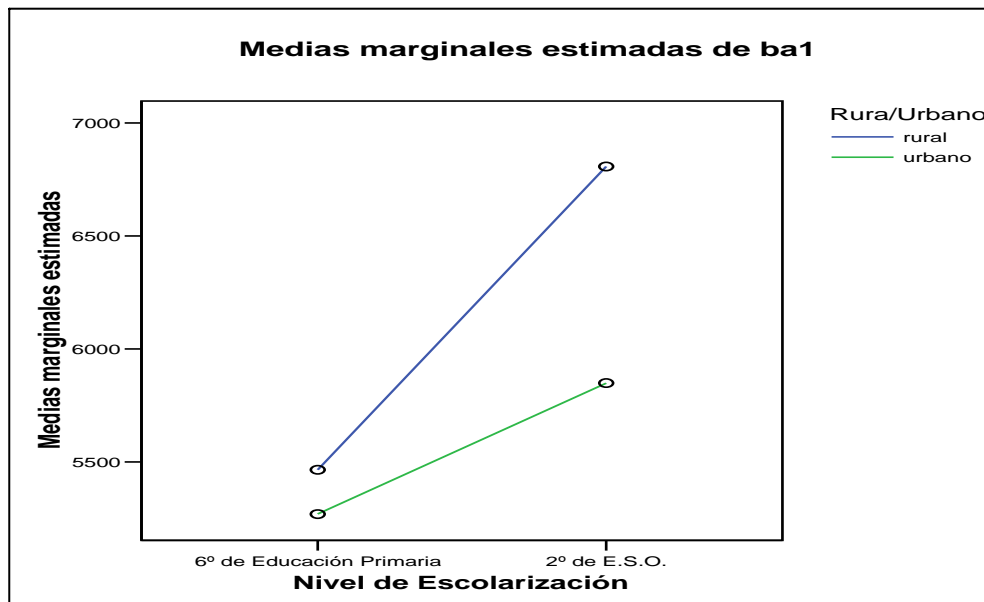


Figura 5.13.: Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Batería 1)

Táboa 5.36.: Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Batería 2)

Pruebas de los efectos inter-sujetos

Variable dependiente: bat2

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	202505059 ^a	3	67501686	8,913	,000	,045
Intersección	2,109E+10	1	2,11E+10	2785,277	,000	,830
nivelesc	138806412	1	1,39E+08	18,328	,000	,031
habitat1	52647111,0	1	52647111	6,952	,009	,012
nivelesc * habitat1	102270,568	1	102270,568	,014	,908	,000
Error	4324404768	571	7573388,4			
Total	2,682E+10	575				
Total corregida	4526909827	574				

a. R cuadrado = ,045 (R cuadrado corregida = ,040)

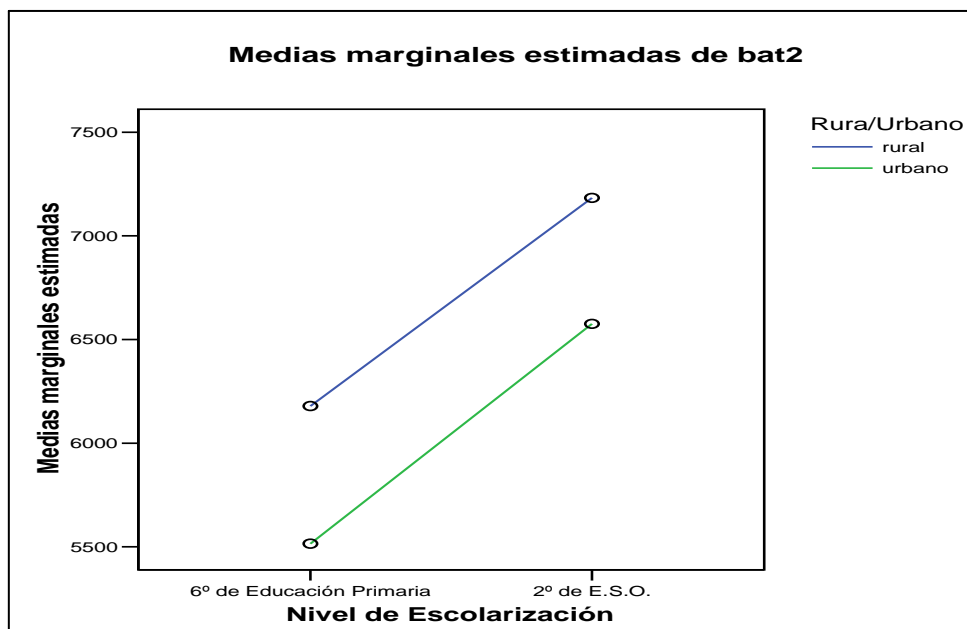


Figura 5.14.: Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano (Batería 2)

Táboa 5.37.: Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano
(Total Test)

Pruebas de los efectos inter-sujetos

Variable dependiente: test

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	171984593 ^a	3	57328198	9,261	,000	,046
Intersección	1,942E+10	1	1,94E+10	3136,792	,000	,846
nivelesc	129386343	1	1,29E+08	20,901	,000	,035
habitat1	47856807,4	1	47856807	7,731	,006	,013
nivelesc * habitat1	4071067,804	1	4071067,8	,658	,418	,001
Error	3534724879	571	6190411,3			
Total	2,423E+10	575				
Total corregida	3706709472	574				

a. R cuadrado = ,046 (R cuadrado corregida = ,041)

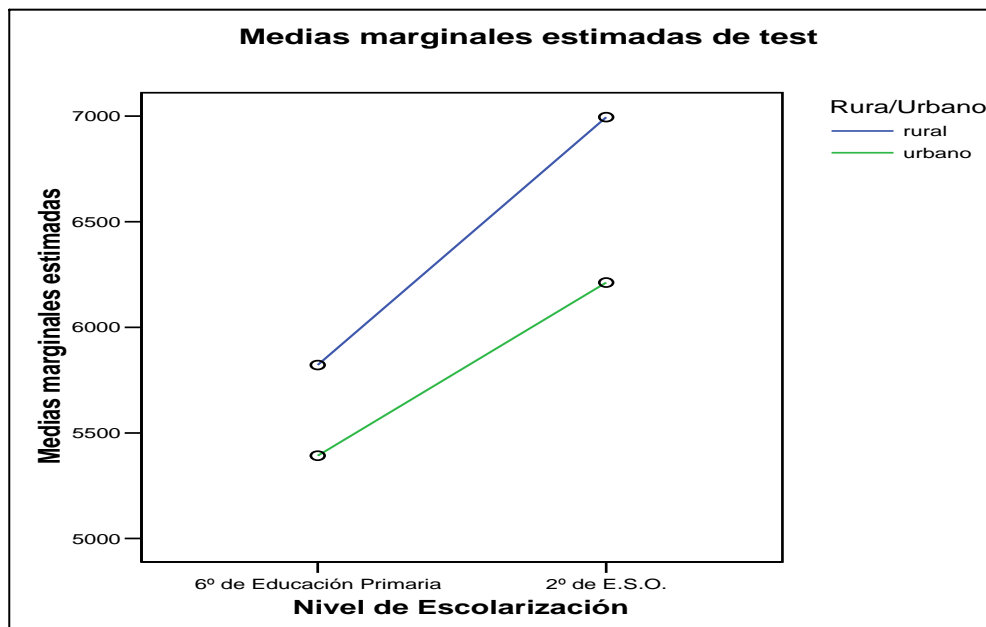


Figura 5.15.: Efectos inter-suxeitos en nivel de escolarización e hábitat rural vs urbano
(Total Test)

Tanto no caso de *6.º de Educación Primaria* coma no de *2.º de ESO* tamén son maiores as medias estimadas do vocabulario correspondentes ós escolares procedentes do ámbito rural que as medias dos que proceden do medio urbano (Táboas 6.3. e 6.4.). A partir da observación dos trazados das figuras 5.13., 5.14. e 5.15., que representan as estimacións do número de palabras coñecidas, é posible efectuar unha interpretación sobre a tendencia á significación das diferenzas entre o alumnado rural e urbano tanto de *6.º de Educación Primaria* coma de *2.º de ESO*. No caso de *2.º de ESO* as liñas tenden a separarse, agás na Batería 2, polo que pode interpretarse que as diferenzas parecen mostrar unha tendencia cara a significación. En *6.º de Educación Primaria* as diferenzas parecen acercarse á significación unicamente na estimación efectuada coa Batería 2, xa que na Batería 1 e no Total do Test as liñas tenden a xuntarse.

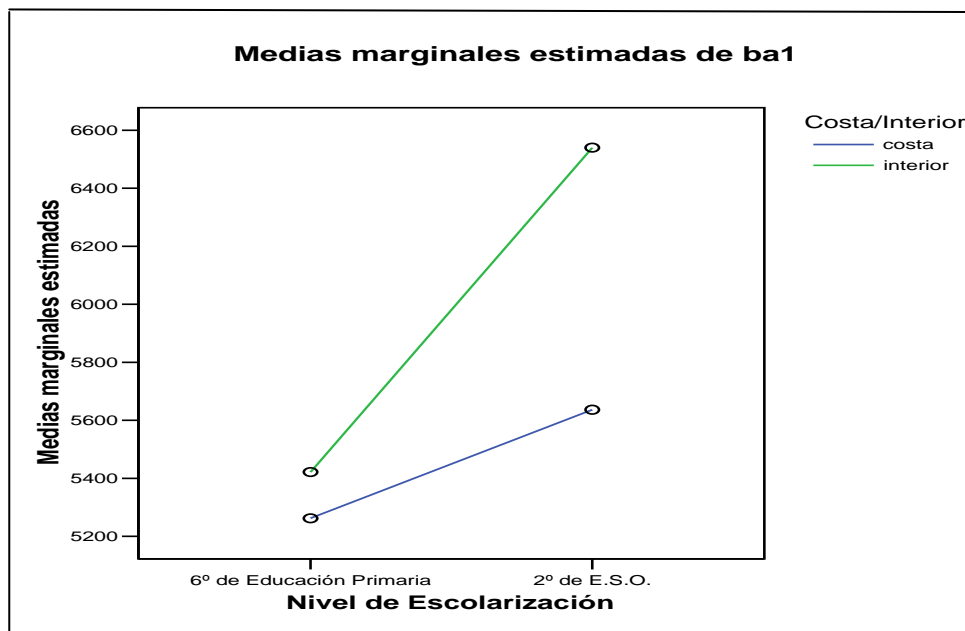
Polo que atinxe á **terceira hipótese**, relativa ás diferenzas entre o alumnado de costa e de interior, no *total da mostra*, que inclúe o alumnado de *6.º de Educación Primaria* e o de *2.º de ESO*, son maiores as medias estimadas do número de palabras coñecidas no caso dos escolares de interior (Táboa 6.5.). As diferenzas no tamaño do vocabulario son significativas (Táboas 5.38., 5.39. e 5.40.), co cal pódese seguir mantendo esta hipótese (Batería 1: $F_{(1, 571)} = 6,346$; $p = 0,012 < 0,05$. Batería 2: $F_{(1, 571)} = 7,448$; $p = 0,007 < 0,05$. Total Test: $F_{(1, 571)} = 7,695$; $p = 0,006 < 0,05$).

Táboa 5.38.: Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Batería 1)**Pruebas de los efectos inter-sujetos**

Variable dependiente: ba1

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	152264166 ^a	3	50754722	8,164	,000	,041
Intersección	1,826E+10	1	1,83E+10	2936,689	,000	,837
nivelesc	77843666,2	1	77843666	12,521	,000	,021
habitat2	39450189,3	1	39450189	6,346	,012	,011
nivelesc * habitat2	19321659,9	1	19321660	3,108	,078	,005
Error	3549897556	571	6216983,5			
Total	2,252E+10	575				
Total corregida	3702161722	574				

a. R cuadrado = ,041 (R cuadrado corregida = ,036)

**Figura 5.16.:** Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Batería 1)

Táboa 5.39.: Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Batería 2)

Pruebas de los efectos inter-sujetos

Variable dependiente: bat2

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	223516424 ^a	3	74505475	9,886	,000	,049
Intersección	2,163E+10	1	2,16E+10	2869,999	,000	,834
nivelesc	118137141	1	1,18E+08	15,675	,000	,027
habitat2	56130879,4	1	56130879	7,448	,007	,013
nivelesc * habitat2	26097770,1	1	26097770	3,463	,063	,006
Error	4303393403	571	7536590,9			
Total	2,682E+10	575				
Total corregida	4526909827	574				

a. R cuadrado = ,049 (R cuadrado corregida = ,044)

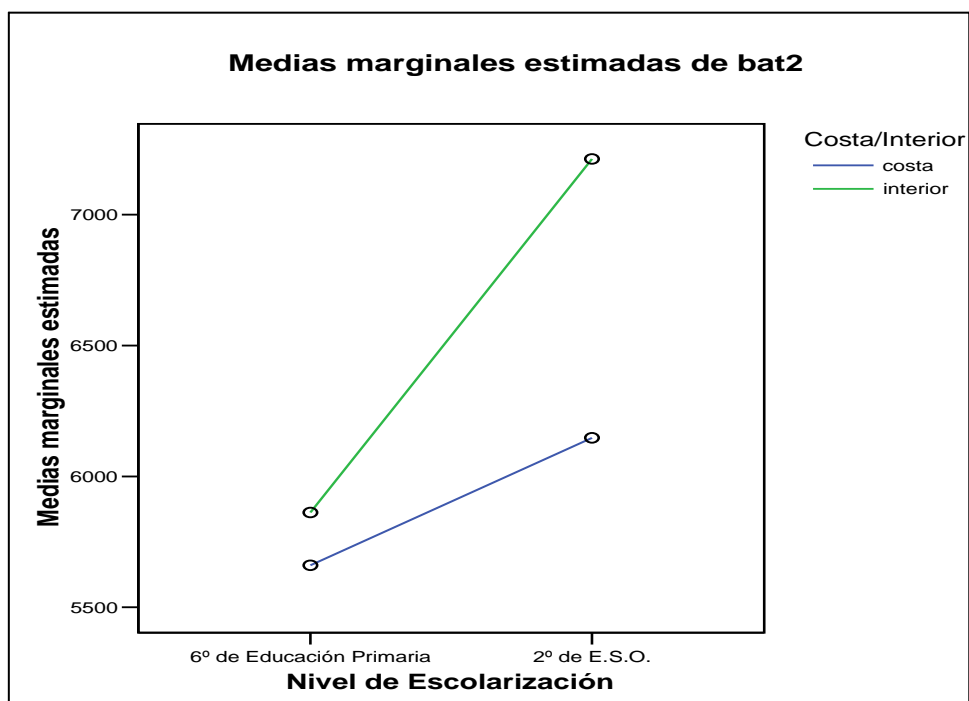


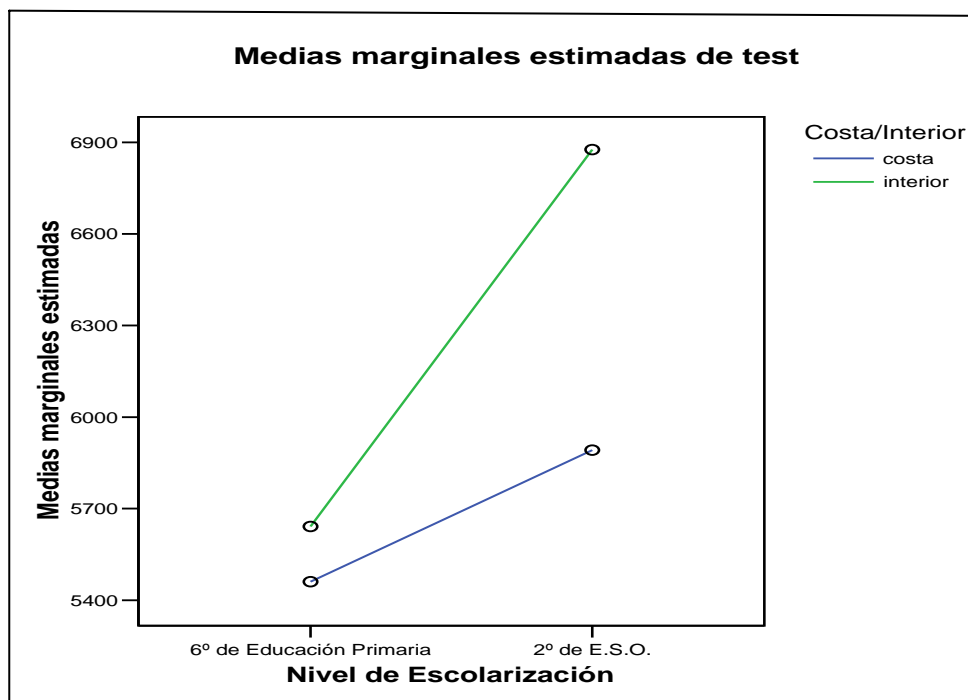
Figura 5.17.: Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Batería 2)

Táboa 5.40.: Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Total Test)**Pruebas de los efectos inter-sujetos**

Variable dependiente: test

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	186238349 ^a	3	62079450	10,069	,000	,050
Intersección	1,991E+10	1	1,99E+10	3228,697	,000	,850
nivelesc	96991258,4	1	96991258	15,731	,000	,027
habitat2	47443272,2	1	47443272	7,695	,006	,013
nivelesc * habitat2	22576421,7	1	22576422	3,662	,056	,006
Error	3520471123	571	6165448,6			
Total	2,423E+10	575				
Total corregida	3706709472	574				

a. R cuadrado = ,050 (R cuadrado corregida = ,045)

**Figura 5.18.:** Efectos inter-suxeitos en nivel de escolarización e hábitat costa vs interior (Total Test)

Respecto da mostra de *6.º de Educación Primaria* (Táboa 6.6.), son lixeiramente maiores as medias estimadas de vocabulario dos escolares de zona de interior, mentres que na mostra de *2.º de ESO* estas diferenzas de medias increméntanse (Táboa 6.7.). Os trazados que aparecen nas Figuras 5.16., 5.17. e 5.18., na medida en que se separan claramente as liñas que representan o número de palabras coñecidas, permiten considerar que en *2.º de ESO* as diferenzas tenden a ser significativas. En *6.º de Educación Primaria*, en cambio, as liñas aproxímanse ata cruzarse practicamente, polo que se pode considerar que non se produce unha tendencia das diferenzas cara á significación.

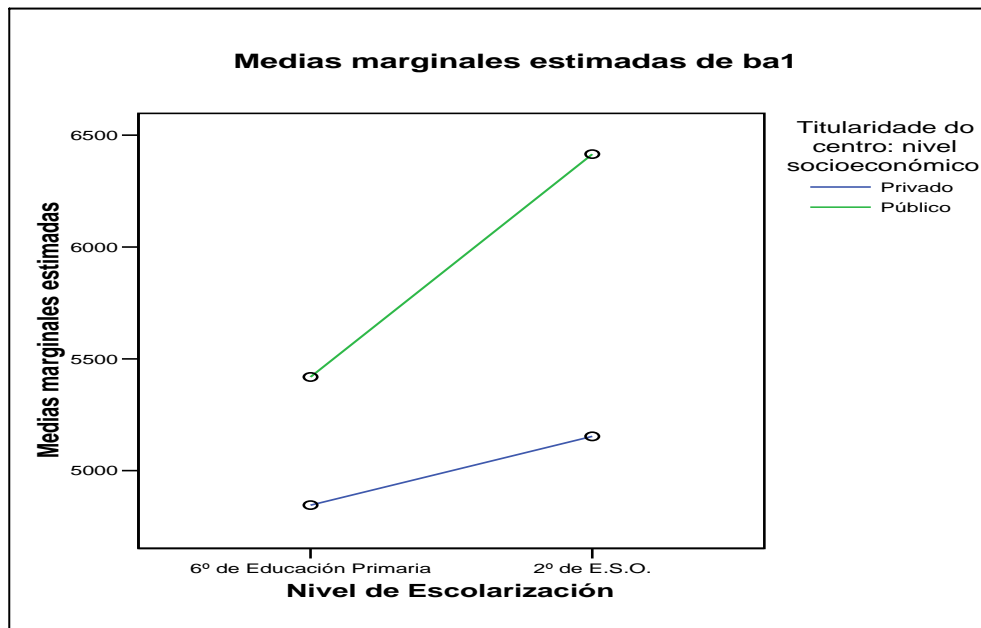
Respecto á **cuarta hipótese**, referida ó agrupamento segundo a titularidade do centro, que serviu de base para a diferenciación en canto ó nivel socioeconómico, para o *total da mostra* (suxeitos de *6.º de Educación Primaria* e de *2.º de ESO*), a media do número de palabras coñecidas polos escolares pertencentes a centros públicos é sempre superior (Táboa 6.8.). As diferenzas (Táboas 5.41., 5.42. e 5.43.) tamén son sempre significativas (Batería 1: $F_{(1, 571)} = 10,444$; $p = 0,001 < 0,05$). Batería 2: $F_{(1, 571)} = 4,016$; $p = 0,046 < 0,05$. Total do Test: $F_{(1, 571)} = 7,423$; $p = 0,007 < 0,05$).

Táboa 5.41.: Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Batería 1)**Pruebas de los efectos inter-sujetos**

Variable dependiente: ba1

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	175191800 ^a	3	58397267	9,454	,000	,047
Intersección	9129262377	1	9,13E+09	1477,985	,000	,721
nivelesc	32541481,5	1	32541481	5,268	,022	,009
titulcen	64513357,4	1	64513357	10,444	,001	,018
nivelesc * titulcen	9089526,311	1	9089526,3	1,472	,226	,003
Error	3526969922	571	6176830,0			
Total	2,252E+10	575				
Total corregida	3702161722	574				

a. R cuadrado = ,047 (R cuadrado corregida = ,042)

**Figura 5.19.:** Efectos inter-suxeitos en nivel de escolarización e titularidade do centro (Batería 1)

Táboa 5.42.: Efectos inter-suxeitos en nivel de escolarización e titularidade do centro
(Batería 2)

Pruebas de los efectos inter-sujetos

Variable dependiente: bat2

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	179271962 ^a	3	59757321	7,848	,000	,040
Intersección	1,124E+10	1	1,12E+10	1476,219	,000	,721
nivelesc	92681659,8	1	92681660	12,172	,001	,021
titulcen	30575761,2	1	30575761	4,016	,046	,007
nivelesc * titulcen	315460,759	1	315460,759	,041	,839	,000
Error	4347637865	571	7614076,8			
Total	2,682E+10	575				
Total corregida	4526909827	574				

a. R cuadrado = ,040 (R cuadrado corregida = ,035)

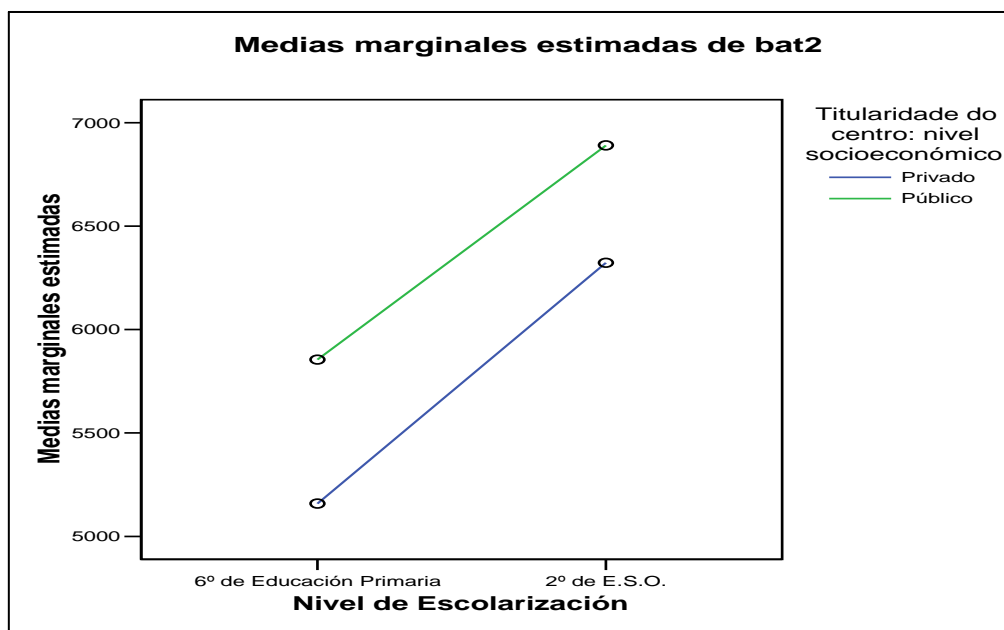


Figura 5.20.: Efectos inter-suxeitos en nivel de escolarización e titularidade do centro
(Batería 2)

Táboa 5.43.: Efectos inter-suxeitos en nivel de escolarización e titularidade do centro
(Total Test)

Pruebas de los efectos inter-sujetos

Variable dependiente: test

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	170992644 ^a	3	56997548	9,205	,000	,046
Intersección	1,016E+10	1	1,02E+10	1640,226	,000	,742
nivelesc	58789136,4	1	58789136	9,494	,002	,016
titulcen	45967233,2	1	45967233	7,423	,007	,013
nivelesc * titulcen	1505327,057	1	1505327,1	,243	,622	,000
Error	3535716829	571	6192148,6			
Total	2,423E+10	575				
Total corregida	3706709472	574				

a. R cuadrado = ,046 (R cuadrado corregida = ,041)

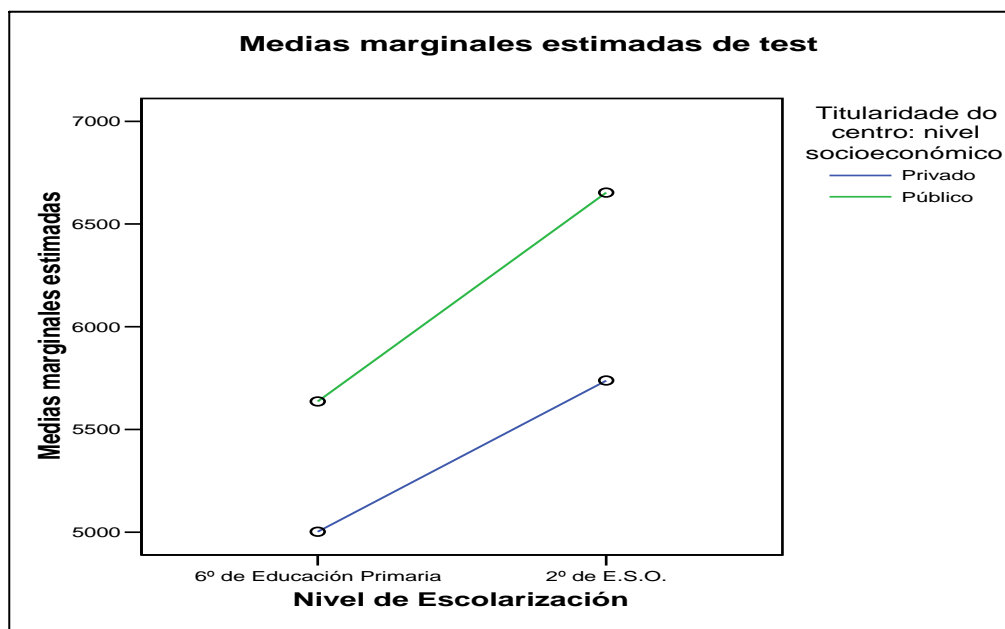


Figura 5.21.: Efectos inter-suxeitos en nivel de escolarización e titularidade do centro
(Total Test)

No relativo á mostra de *6.º de Educación Primaria* (Táboa 6.9.), as medias estimadas do vocabulario son superiores para os escolares que proceden dos centros públicos. Se se observa a dirección das liñas que representan o tamaño do vocabulario (Figuras 5.19., 5.20. e 5.21.), non parece que se poda deducir que o tamaño do vocabulario do alumnado de centros públicos tenda a ser realmente maior que o do alumnado de centros privados.

En canto á mostra de *2.º de ESO* (Táboa 6.10.), novamente son superiores as medias correspondentes ó alumnado procedente dos centros públicos. Neste caso, en xeral, a dirección das liñas que representan o vocabulario nas Figuras 5.19., 5.20. e 5.21. tenden a separarse, polo que si se pode pensar que as diferenzas apuntan á significación.

No que atinxe á **quinta hipótese** (diferenzas segundo a lingua familiar habitual), para o *total da mostra* (Táboas 6.11. e 6.12.), as medias da estimación do número de palabras coñecidas son superiores no caso dos escolares que teñen o galego como lingua habitual. Como se pode comprobar nas Táboas 5.44., 5.46. e 5.48., as diferenzas son significativas (Batería 1: $F_{(2, 569)} = 12,043$; $p = 0,000 < 0,05$. Batería 2: $F_{(2, 569)} = 12,619$; $p = 0,000 < 0,05$. Total do Test: $F_{(2, 569)} = 13,617$; $p = 0,000 < 0,05$). Unha análise post hoc mediante a proba *Scheffé* (Táboas 5.45., 5.47. e 5.49.) mostra que se producen diferenzas significativas entre os escolares que teñen o galego como lingua habitual e os que usan o castelán (Batería 1: $p = 0,000 < 0,05$. Batería 2: $p = 0,000 < 0,05$. Total do Test: $p = 0,000 < 0,05$), así como entre os escolares que utilizan o galego de forma habitual e os que utilizan as dúas (Batería 1: $p = 0,002 < 0,05$). Batería 2: $p = 0,000 < 0,05$. Total do Test: $p = 0,000 < 0,05$). Por conseguinte, pódese considerar que os datos permiten mante-la hipótese no sentido de que é significativamente maior o vocabulario do alumnado que usa habitualmente o galego.

Táboa 5.44.: Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Batería 1)**Pruebas de los efectos inter-sujetos**

Variable dependiente: ba1

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	258262030 ^a	5	51652406	8,534	,000	,070
Intersección nivelesc	1,823E+10	1	1,82E+10	3012,728	,000	,841
Lingua	104107739	1	1,04E+08	17,201	,000	,029
nivelesc * Lingua	145780290	2	72890145	12,043	,000	,041
Error	11285828,9	2	5642914,5	,932	,394	,003
Total	3443899692	569	6052547,8			
Total corregida	2,252E+10	575				
Total corregida	3702161722	574				

a. R cuadrado = ,070 (R cuadrado corregida = ,062)

Táboa 5.45.: Comparacións post hoc en nivel de escolarización e lingua habitual (Batería 1)**Comparaciones múltiples**

Variable dependiente: ba1

Scheffe

(I) Lingua	(J) Lingua	Diferencia entre medias (I-J)	Error típ.	Significación	Intervalo de confianza al 95%.	
					Límite inferior	Límite superior
galego	castelán	1246,67*	253,511	,000	624,50	1868,84
	as dúas	972,39*	269,581	,002	310,78	1633,99
castelán	galego	-1246,67*	253,511	,000	-1868,84	-624,50
	as dúas	-274,28	242,578	,528	-869,62	321,05
as dúas	galego	-972,39*	269,581	,002	-1633,99	-310,78
	castelán	274,28	242,578	,528	-321,05	869,62

Basado en las medias observadas.

*. La diferencia de medias es significativa al nivel ,05.

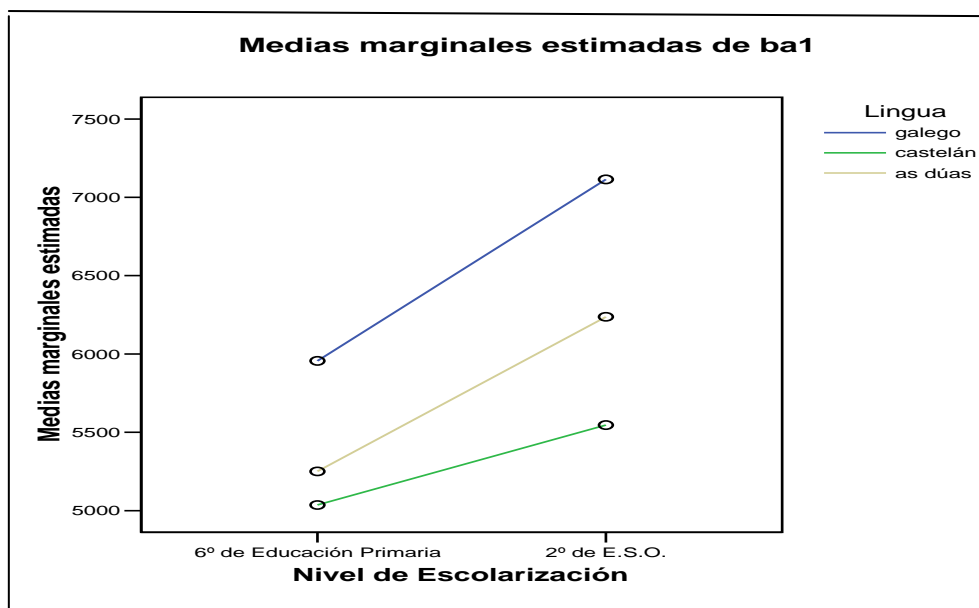


Figura 5.22.: Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Batería 1)

Táboa 5.46.: Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Batería 2)

Pruebas de los efectos inter-sujetos

Variable dependiente: bat2

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	335339540 ^a	5	67067908	9,104	,000	,074
Intersección	2,144E+10	1	2,14E+10	2909,808	,000	,836
nivelesc	123951295	1	1,24E+08	16,826	,000	,029
Lingua	185923078	2	92961539	12,619	,000	,042
nivelesc * Lingua	2863231,035	2	1431615,5	,194	,823	,001
Error	4191570287	569	7366555,9			
Total	2,682E+10	575				
Total corregida	4526909827	574				

a. R cuadrado = ,074 (R cuadrado corregida = ,066)

Táboa 5.47.: Comparacións post hoc en nivel de escolarización e lingua habitual
(Batería 2)

Comparaciones múltiples

Variable dependiente: bat2

Scheffe

(I) Lingua	(J) Lingua	Diferencia entre medias (I-J)	Error típ.	Significación	Intervalo de confianza al 95%.	
					Límite inferior	Límite superior
galego	castelán	1360,87*	279,679	,000	674,49	2047,26
	as dúas	1262,20*	297,408	,000	532,30	1992,10
castelán	galego	-1360,87*	279,679	,000	-2047,26	-674,49
	as dúas	-98,68	267,618	,934	-755,47	558,11
as dúas	galego	-1262,20*	297,408	,000	-1992,10	-532,30
	castelán	98,68	267,618	,934	-558,11	755,47

Basado en las medias observadas.

*. La diferencia de medias es significativa al nivel ,05.

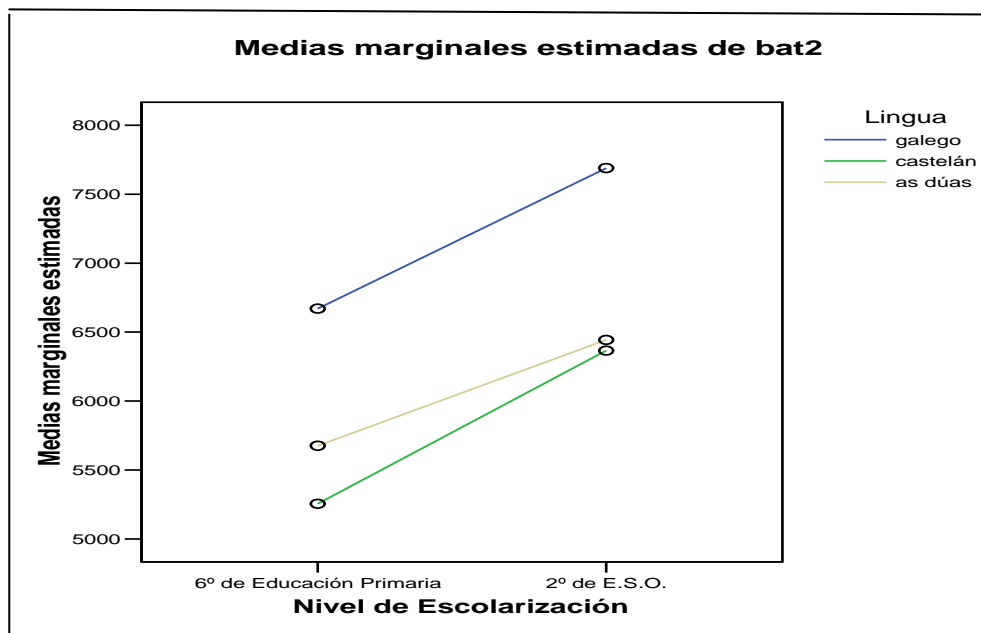


Figura 5.23.: Efectos inter-suxeitos en nivel de escolarización e lingua habitual
(Batería 2)

Táboa 5.48.: Efectos inter-suxeitos en nivel de escolarización e lingua habitual
(Total Test)

Pruebas de los efectos inter-sujetos

Variable dependiente: test

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	288912345 ^a	5	57782469	9,620	,000	,078
Intersección nivelesc	1,980E+10	1	1,98E+10	3296,484	,000	,853
Lingua	113882915	1	1,14E+08	18,959	,000	,032
nivelesc * Lingua	163589861	2	81794930	13,617	,000	,046
Error	1891331,971	2	945665,986	,157	,854	,001
Total	3417797128	569	6006673,3			
Total corregida	2,423E+10	575				
Total corregida	3706709472	574				

a. R cuadrado = ,078 (R cuadrado corregida = ,070)

Táboa 5.49.: Comparacións post hoc en nivel de escolarización e lingua habitual
(Total Test)

Comparaciones múltiples

Variable dependiente: test

Scheffe

(I) Lingua	(J) Lingua	Diferencia entre medias (I-J)	Error típ.	Significación	Intervalo de confianza al 95%.	
					Límite inferior	Límite superior
galego	castelán	1303,46*	252,548	,000	683,66	1923,26
	as dúas	1116,91*	268,558	,000	457,81	1776,00
castelán	galego	-1303,46*	252,548	,000	-1923,26	-683,66
	as dúas	-186,55	241,657	,742	-779,63	406,52
as dúas	galego	-1116,91*	268,558	,000	-1776,00	-457,81
	castelán	186,55	241,657	,742	-406,52	779,63

Basado en las medias observadas.

*. La diferencia de medias es significativa al nivel ,05.

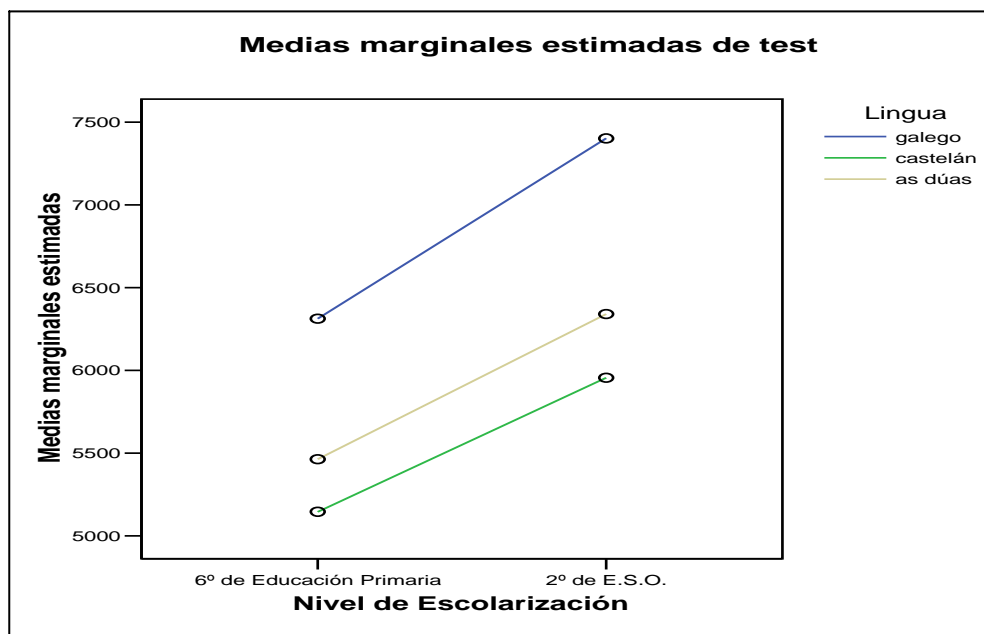


Figura 5.24.: Efectos inter-suxeitos en nivel de escolarización e lingua habitual (Total Test)

No relativo ás mostras de *6.º de Educación Primaria* e de *2.º de ESO*, as medias do vocabulario estimado dos escolares que falan habitualmente en galego superan ás medias dos que habitualmente o fan en castelán e ás medias dos que utilizan as dúas linguas (Táboas 6.13. a 6.16.). A través da observación das Figuras 5.22., 5.23. e 5.24., en base ós trazados das liñas que representan o tamaño do vocabulario (tenden a non cruzarse), as diferenzas parecen ser claras entre o alumnado que utiliza o galego en comparación co que fala tanto castelán coma co que fala as dúas linguas.

En termos xerais (Táboa 5.50., p. 295), a partir da análise dos datos comentados sobre a validez de construto, pódese dicir que hai diferenzas significativas no tamaño do vocabulario en función da idade, o que é o esperable. De non ter detectado estas diferenzas, o test carecería de utilidade.

Tamén se poden manter as outras catro hipóteses, tanto se se consideran as Baterías separadamente coma se se consideran como metades dun só test, é dicir, existen diferenzas significativas no tamaño do vocabulario, correlativamente, entre o alumnado do hábitat rural e urbano, entre o de hábitat costeiro e de interior, entre o escolarizado en centros públicos e privados, así como entre o que fala habitualmente en galego e o que fala en castelán ou nas dúas linguas.

En relación coa análise baseada nos trazados gráficos dos resultados en cada un dos niveis de idade, sen perder de vista que non foron significativas as respectivas interaccións ó nivel do 5% nas ANOVAS, á vez que tendo en conta a necesidade dunha replicación dos datos que se presentan neste traballo, obsérvase que as diferenzas tenden a desaparecer a medida que se descende en idade, fundamentalmente nos hábitat rural/urbano e costa/interior. Pero tamén se produce esta tendencia en función da titularidade do centro, xa que as diferenzas parecen ser claras en 2.º de ESO, mentres que non ocorre o mesmo en 6.º de Educación Primaria. Unha posible explicación desta tendencia sería a existencia do proceso mediante o cal o galego deixa de ser progresivamente a lingua familiar habitual, co que cada vez máis é o ámbito escolar no que o alumnado aprende esta lingua e, con elo, exerce un efecto igualador na competencia nesta lingua. Unha segunda explicación, sería que, do mesmo xeito que a medida que se incrementa a idade e o nivel educativo, se producen unhas diferenzas individuais maiores na calidade e cantidade de coñecemento en xeral entre a poboación escolar dunha mesma idade, tamén cabe a posibilidade de que, coa idade e nivel escolar, se acrecenten as diferenzas en vocabulario dentro da idade e nivel. Pola súa parte, entre os escolares co galego como lingua habitual no contexto familiar e os que usan habitualmente o castelán ou as dúas linguas, as diferenzas parecen ser claras tanto en 6.º de Educación Primaria coma en 2.º de ESO.

Táboa 5.50.: Resume da significación das diferenzas entre os distintos agrupamentos

Factores	Mostra Total		
	Bat. 1	Bat. 2	Test
Nivel de escolarización	**	**	**
Rural/Urbano	**	**	**
Costa/Interior	**	**	**
Nivel socioeconómico	**	*	**
Lingua familiar:			
- galego/castelán	**	**	**
- galego/as dúas	**	**	**
- castelán/as dúas	--	--	--

** significación ó 0,01 * significación ó 0,05

5.5.3 Outros tipos de validez

A **validez de contido** queda asegurada polo proceso de construción dos subtests descrito no apartado 3 do Capítulo 4, que garante que as palabras que os compoñen constitúen unha mostra representativa das que contén o dicionario da RAG, o contido ou dominio que se avalía neste caso.

Efectuouse, así mesmo, unha análise do grao de facilidade/dificultade de tódolos ítems sen empregar ningún procedemento estatístico, co obxecto de achegar datos que poideran apoiar a validez de contido do instrumento. A análise consistiu en verificar a través dunha simple inspección se se producía unha correspondencia, en termos de porcentaxes, entre o número de escolares que sinalaron as distintas palabras e os niveis de facilidade/dificultade establecidos previamente para as palabras por medio da valoración de tres docentes de ensino non universitario. O grao de dificultade das pseudopalabras

enténdese como análogo ó das palabras de nivel 4, as cales pódense considerar como infrecuentes e moi pouco familiares. Se existe congruencia co xuízo acerca da dificultade das palabras emitido polos profesores con experiencia, contarase cun elemento máis no que sustenta-la validez.

No Anexo VIII recóllense a modo de exemplificación as porcentaxes de resposta ós 60 ítems de 4 subtest ó azar (1 e 3 da Batería 1; 3 e 5 da Batería 2), para a mostra Total (N=575), a de sexto de Educación Primaria (N=313) e a de segundo de ESO (N=262).

Na última columna das táboas figuran as 20 pseudopalabras do subtest e á dereita a porcentaxe de suxeitos que as sinalaron coma se fosen palabras reais. As seleccionadas nunha maior porcentaxe son: *esbarecer* no subtest 1.1 en tódalas mostras (55% na mostra total, 57% na mostra de sexto de Educación Primaria e 53% na mostra de segundo de ESO); *tragalinguas* no subtest 1.3 tamén en tódalas mostras (44%, 43% e 45%); *conversatorio* no subtest 2.3 para a mostra Total (34%) e a de sexto de Educación Primaria (37%), mentres que en segundo de ESO as máis sinaladas neste subtest son *tintaxe* e *conversatorio* nunha porcentaxe practicamente igual (33% e 34%, respectivamente); por último, no subtest 2.5 a máis elixida é *cerraxeiro*, que é sinalada nas seguintes porcentaxes, nas mostras total, de 6.º de Educación Primaria e de 2.º de ESO de forma correlativa: 73%, 69% e 78%. O feito de que estes ítems, xunto con outros, sexan sinalados en porcentaxes considerables, pode deberse nuns casos, como *esbarecer* ou *cerraxeiro*, a que son moi semellantes a palabras reais ou, incluso, como sucede co segundo, son utilizados habitualmente de forma inadecuada, mentres que noutros casos existe a posibilidade de que se cometan erros na decodificación e se lesen como palabras reais, como en *tragalinguas* ou *conversatorio*. En todo caso, a porcentaxe media de veces en que se sinalan este tipo de ítems en conxunto é reducida, e en ningún dos subtest nin en ningunha das mostras supera o 14%.

Nas primeiras catro columnas, están situadas as palabras dos catro niveis de dificultade, establecidos en base á valoración do profesorado. Como se pode comprobar

(Anexo VIII), consultando na parte inferior (última fila) de cada unha das táboas, as medias das porcentaxes de palabras do nivel 1 sinaladas son substancialmente maiores que as do resto dos niveis, sendo as do nivel 4, ademais, moi inferiores as dos niveis 2 e 3. En ningún subtest nin en ningunha mostra superan o 13%, de modo que estas palabras do nivel 4, que o profesorado considerou como as máis difíciles, son sinaladas en porcentaxes semellantes ós das pseudopalabras.

Parece, polo tanto, que a categorización das palabras efectuada polos xuíces atendendo ó grao de dificultade que representan na súa opinión, para, posteriormente, incluílas en función deste grao de dificultade nos subtest, correspóndese coa dificultade que na práctica representan para os escolares ós que se lle administraron, o que se pode entender como un dato máis en apoio da validez do test.

En canto á **validez empírica ou relacionada cun criterio**, atendendo o deseño utilizado nesta investigación para calcula-lo coeficiente, correspóndese coa denominada máis concretamente **validez concorrente**. O procedemento consistiu en relaciona-lo rendemento no test coa valoración do rendemento xeral do alumnado na materia de lingua galega por parte do profesorado que a impartía. Pedíuselle a estes docentes que cualificaran o rendemento como *Baixa*, *Medio* ou *Alto*. Polo procedemento utilizado, incluso se pode entender esta validez como de tipo retrospectivo, consonte sinala Muñiz (2000, p. 153).

Debe de terse en conta que o rendemento avalíase utilizando o coñecemento que demostra o alumnado sobre diferentes contidos verbais, entre os que o vocabulario non é máis que un deles e seguramente tampouco é ó que se lle atribúe un maior peso directo na cualificación da materia. Pero, indirectamente, pode constituír un elemento que favoreza a adquisición doutros coñecementos, así como a súa exposición. Sendo conscientes deste feito e das limitacións que implica, en tanto que non se dispón doutro criterio que poda ser utilizado para a validación, decidiuse facer uso deste.

As correlacións entre a estimación do número de palabras coñecidas e o rendemento xeral en lingua galega para o *total da mostra*, utilizando o coeficiente de correlación de Pearson, están comprendidas entre 0,509 para a Batería 1 e 0,512 para a Batería 2, ata 0,537 co total do Test (Táboa 5.51.).

Táboa 5.51.: Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra total (N=575)

	Batería 1	Batería 2	Total Test
Rendemento	0,509**	0,512**	0,537**

** A correlación é significativa ó 0,01% (bilateral).

Os valores son obviamente inferiores (Táboa 5.52.) cando se trata das correlacións entre as puntuacións nos diferentes subtests e o rendemento. Sitúanse en torno a 0,30, correspondéndolle a máis baixa ó subtest 1.º (0,247) e a máis alta ó subtest 9.º (0,402).

Táboa 5.52.: Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra total (N=575)

Subtests	1	2	3	4	5	6	7	8	9	10
Rendemento	,247**	,312**	,371**	,348**	,327**	,316**	,389**	,378**	,402**	,313**

** A correlación é significativa ó 0,01% (bilateral).

En canto á mostra de *6.º de Educación Primaria*, os índices de correlación obtidos entre o vocabulario estimado e o rendemento xeral en lingua galega, abarcan desde 0,548 coa Batería 1 a 0,535 coa Batería 2 e 0,571 co total do Test (Táboa 5.53.).

Táboa 5.53.: Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 6.º de E. Primaria (N=313)

	Batería 1	Batería 2	Total Test
Rendemento	0,548**	0,535**	0,571**

** A correlación é significativa ó 0,01% (bilateral).

En canto ás correlacións cos 10 subtests (Táboa 5.54.), os valores están na súa meirande parte por encima de 0,30, agás co subtest 1.º (0,187), sendo a máis alta a que lle corresponde ó subtest 8.º (0,440).

Táboa 5.54.: Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 6.º de E. Primaria (N=313)

Subtests	1	2	3	4	5	6	7	8	9	10
Rendemento	,187**	,313**	,411**	,407**	,351**	,307**	,376**	,440**	,395**	,298**

** A correlación é significativa ó 0,01% (bilateral).

Para a mostra de 2.º de ESO, as correlacións entre o número estimado de palabras e a estimación do rendemento en galego efectuada polo profesorado, sitúase entre 0,461 para a Batería 1 e 0,487 para a Batería 2, ata 0,500 co total do Test (Táboa 5.55.).

Táboa 5.55.: Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 2.º de ESO (N=262)

	Batería 1	Batería 2	Total Test
Rendemento	0,461**	0,487**	0,500**

** A correlación é significativa ó 0,01% (bilateral).

No relativo ás correlacións cos 10 subtests (Táboa 5.56), agás co 4.º (0,278) e co 8.º (0,296), sitúanse por encima de 0,30, sendo a máis alta a que se produce co subtest 9.º (0,408).

Táboa 5.56.: Correlacións entre a estimación do número de palabras coñecidas e o rendemento en galego na mostra de 2.º de ESO (N=262)

Subtests	1	2	3	4	5	6	7	8	9	10
Rendemento	,315**	,302**	,315**	,278**	,298**	,328**	,407**	,296**	,408**	,329**

** A correlación é significativa ó 0,01% (bilateral).

Polo tanto, en resumo, se se consideran os datos obtidos a partir da estimación do número de palabras coñecidas coas Baterías 1 e 2 e co total do Test, os coeficientes, aínda sendo baixos, acércanse a valores aceptables (Downie e Heath, 1983; Rust e Golombok, 1995). Ademais, débese ter en conta que a finalidade deste test non é predici-la posición dun determinado suxeito na distribución da variable criterio (Anastasi, 1977; Cohen e Manion, 1990).

Como síntese, no que se refire á validez do test, en primeiro lugar, os resultados da comprobación de hipóteses permiten concluír que a validez de construto está avalada pola capacidade da proba para diferenciar distintos grupos de escolares en función do nivel educativo, do hábitat xeográfico, da titularidade do centro de procedencia e da lingua familiar habitual. Tamén se pode apoiar esta validez na achega teórica da primeira parte da tese e na descrición do proceso de constitución da mostra.

En segundo lugar, a validez de contido está asegurada polo proceso de construción do test e pola congruencia que se da entre o grao de dificultade que teñen os ítems en opinión de xuíces docentes e o grao de dificultade dos devanditos ítems obtido en base ás respostas dos escolares no test.

Por último, a validez referida a criterio pode sustentarse nas aceptables correlacións que se obtiveron entre a estimación mediante o test do número de palabras coñecidas e a

valoración do rendemento en lingua galega efectuado polos docentes que impartían esta materia.

Capítulo 6

Elaboración de baremos. Suxestións para a súa utilización

A mellor utilización dos datos conseguidos mediante este tipo de probas é a estimación do número de palabras coñecidas por un escolar en concreto (tamaño do vocabulario dese escolar), aínda que podería estimarse igualmente a dun grupo como media. Como xa se indicou, estes datos obtéñense consultando a Táboa que aparecen no Anexo IV, coas puntuacións de Acertos e de Erros.

Tamén pode resultar de utilidade a comparación da puntuación estimada dun escolar determinado con respecto ás dun grupo normativo. Como é sabido, as puntuacións directas necesitan unha elaboración posterior, unha transformación, para dotalas dunha maior significación. No caso do test construído na presente investigación, a puntuación directa é unha estimación absoluta do vocabulario coñecido (cando menos, a un nivel básico de coñecemento), pero pode adquirir máis significado cando se compara coa estimación do vocabulario coñecido doutros suxeitos.

Neste capítulo, en primeiro lugar, preséntanse algúns datos descritivos ós que se aludiu no apartado sobre validez de construto. Son datos relativos tanto á mostra total coma ás mostras de 6.º de Educación Primaria e de 2.º de ESO, e inclúense aquí co fin de ofrecer unha visión da estimación do tamaño do vocabulario do alumnado pertencente ós diferentes agrupamentos cos que se traballou. En segundo lugar, preséntanse baremos que permiten efectuar unha comparación entre suxeitos, tanto de 6.º de Educación Primaria coma de 2.º de ESO. Agora ben, tódolos datos que se achegan deben interpretarse con cautela, a teor das limitacións da investigación xa sinaladas en canto ás características da mostra (falta de aleatorización, reducido tamaño do número de suxeitos nalgúns agrupamentos, etc.).

6.1. Datos descriptivos básicos

Primeiramente, en función da idade e nivel de escolarización (Táboa 6.1.), os datos indican que a media das puntuacións do alumnado de 2.º de ESO resulta ser superior á do alumnado de 6.º de Educación Primaria nunhas 832 palabras para o cálculo mediante a Batería 1, en 1021 aproximadamente co cálculo mediante a Batería 2 e en torno a 927 utilizando o Total do Test para a estimación.

Táboa 6.1.: Estatísticos descriptivos da estimación do vocabulario nas mostras de 6.º de Educación Primaria e 2.º de ESO

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 6º de E. Primaria	313	206	10.826	5342,17	2380,21	134,54
	262	618	12.991	6174,45	2652,09	163,85
Batería 2 6º de E. Primaria	313	206	12.373	5761,20	2595,69	146,72
	262	618	14.538	6782,04	2953,12	182,44
Total Test 6º de E. Primaria	313	206	11.187	5551,35	2355,79	133,16
	262	618	13.507	6478,11	2664,28	164,60

En segundo lugar, en canto á primeira distinción segundo o hábitat, a media da estimación de palabras coñecidas do grupo de poboación rural é superior en tódalas comparacións, como se pode observar nunha primeira inspección (Táboa 6.2.). No relativo ó total da mostra, as diferenzas apenas superan as 600 palabras na Batería 2, que é na que se produce a maior diferenza. No caso da mostra de 6.º de Educación Primaria (Táboa 6.3.), o resultado é semellante. Non obstante, na mostra de 2.º de ESO as diferenzas oscilan entre máis de 900 palabras na Batería 1 e arredor de 600 na Batería 2, sempre a favor do grupo rural (Táboa 6.4.).

Táboa 6.2.: Estatísticos descritivos da estimación do vocabulario na mostra total para os ámbitos rural e urbano

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	rural	205	618	11.445	6047,90	2421,28	169,11
	urbano	370	206	12.991	5540,50	2588,40	134,56
Batería 2	rural	205	618	12.476	6615,04	2610,72	182,34
	urbano	370	206	14.538	6010,99	2893,01	150,40
Total Test	rural	205	824	11.393	6331,08	2368,92	165,45
	urbano	370	206	13.507	5775,58	2613,91	135,89

Táboa 6.3.: Estatísticos descritivos da estimación do vocabulario na mostra de 6.º de Educación Primaria para os ámbitos rural e urbano

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	rural	116	618	10.620	5465,36	2335,97	216,89
	urbano	197	206	10.826	5269,63	2408,83	171,62
Batería 2	rural	116	618	10.826	6179,01	2579,25	239,48
	urbano	197	206	12.373	5515,19	2580,26	183,84
Total Test	rural	116	824	10.001	5821,59	2311,42	214,61
	urbano	197	206	11.187	5392,21	2372,96	169,07

Táboa 6.4.: Estatísticos descritivos da estimación do vocabulario na mostra de 2.º de ESO para os ámbitos rural e urbano

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	rural	89	1031	11.445	6807,17	2330,06	246,99
	urbano	173	618	12.991	5848,95	2753,47	209,34
Batería 2	rural	89	618	12.476	7183,36	2555,25	270,86
	urbano	173	618	14.538	6575,58	3125,17	237,60
Total Test	rural	89	825	11.393	6995,12	2288,91	242,62
	urbano	173	618	13.507	6212,14	2807,40	213,44

En terceiro lugar, no relativo á segunda distinción do hábitat, comparando a estimación do número de palabras coñecidas dos grupos de poboación costeira e de interior, as diferenzas sitúanse a favor do grupo de interior en tódolos casos. Para a mostra total de escolares a media é lixeiramente superior a 500 palabras na Batería 1 e lixeiramente superior a 600 na Batería 2 e no Total do Test (Táboa 6.5.). Para a mostra de 6.º de Educación Primaria (Táboa 6.6.), as diferenzas medias redúcense claramente (están por debaixo de 200 palabras tanto na Batería 1 coma no Total do Test e en 201 na Batería 2). En cambio, na mostra de 2.º de ESO estas diferenzas incrementáanse, de forma que a media do grupo de interior é superior en algo máis de 900 palabras na Batería 1 e en algo máis de 1000 na Batería 2 ou case 1000 no Total do Test (Táboa 6.7.).

Táboa 6.5.: Estatísticos descritivos da estimación do vocabulario na mostra total para os ámbitos costa e interior

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 costa	262	206	11.445	5413,69	2313,07	142,90
interior	313	618	12.991	5978,97	2691,49	152,13
Batería 2 costa	262	206	12.167	5857,22	2487,74	153,69
interior	313	618	14.538	6535,34	3020,25	170,71
Total Test costa	262	206	11.702	5635,13	2243,10	138,58
interior	313	618	13.507	6256,97	2737,64	154,74

Táboa 6.6.: Estatísticos descritivos da estimación do vocabulario na mostra de 6.º de Educación Primaria para os ámbitos costa e interior

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	costa	156	206	10.826	5262,17	2269,75	181,72
	interior	157	618	10.620	5421,66	2489,85	198,71
Batería 2	costa	156	206	10.826	5660,07	2379,72	190,53
	interior	157	618	12.373	5861,69	2797,86	223,29
Total Test	costa	156	206	10.569	5460,69	2172,47	173,94
	interior	157	618	11.187	5641,42	2528,59	201,80

Táboa 6.7.: Estatísticos descritivos da estimación do vocabulario na mostra de 2.º de ESO para os ámbitos costa e interior

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	costa	106	1237	11.445	5636,69	2368,55	230,05
	interior	156	618	12.991	6539,85	2776,89	222,33
Batería 2	costa	106	1237	12.167	6147,37	2623,22	254,79
	interior	156	618	14.538	7213,29	3092,04	247,56
Total Test	costa	106	1391	11.702	5891,86	2329,81	226,29
	interior	156	618	13.507	6876,47	2807,31	224,76

En cuarto lugar, para os agrupamentos segundo a titularidade do centro, que se adoptou como base para establecer unha diferenciación en canto ó nivel socioeconómico do alumnado, a media do número de palabras estimadas coñecidas polos escolares pertencentes á centros públicos é sempre superior. Na mostra total as diferenzas oscilan entre máis de 800 palabras na Batería 1 e pouco máis de 500 na Batería 2 (Táboa 6.8.). Na mostra de 6.º de Educación Primaria, as diferenzas das medias estimadas do vocabulario están entre cerca de 600 palabras na Batería 1 e practicamente 700 na Batería 2 (Táboa 6.9.). Os datos relativos á mostra de 2.º de ESO son realmente dispares, xa que na Batería 1

a diferenza acércase a 1300 palabras, na Batería 2 ás 600 e no Total do Test é de pouco máis de 900 (Táboa 6.10.).

Táboa 6.8.: Estatísticos descritivos da estimación do vocabulario na mostra total segundo a titularidade do centro

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	privado	92	206	11.136	5012,87	2612,10	272,42
	público	483	618	12.991	5856,36	2505,53	114,01
Batería 2	privado	92	206	13.094	5791,32	3111,29	324,37
	público	483	618	14.538	6309,22	2742,51	124,79
Total Test	privado	92	206	11.579	5401,95	2753,01	287,02
	público	483	618	13.507	6082,52	2486,99	113,16

Táboa 6.9.: Estatísticos descritivos da estimación do vocabulario na mostra de 6.º de Educación Primaria segundo a titularidade do centro

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	privado	42	206	9795	4845,86	2378,83	367,06
	público	271	618	10.826	5419,09	2375,53	144,30
Batería 2	privado	42	206	9279	5158,64	2556,52	394,48
	público	271	618	12.373	5854,59	2593,83	157,56
Total Test	privado	42	206	9382	5002,02	2353,34	363,13
	público	271	618	11.187	5636,48	2349,01	142,69

Táboa 6.10.: Estatísticos descritivos da estimación do vocabulario na mostra de 2.º de ESO segundo a titularidade do centro

		N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1	privado	50	618	11.136	5153,16	2810,91	397,52
	público	212	1031	12.991	6415,32	2561,24	175,91
Batería 2	privado	50	618	13.094	6322,76	3446,36	487,39
	público	212	618	14.538	6890,36	2822,66	193,86
Total Test	privado	50	618	11.579	5737,88	3031,36	428,70
	público	212	825	13.507	6652,70	2546,77	174,91

En quinto lugar, en base á distinción conforme á lingua habitual na familia, a media da estimación do número de palabras coñecidas é superior no grupo que ten o galego como lingua habitual. Así, na mostra total superan ós que habitualmente falan en castelán entre 1247 palabras na Batería 1 e 1361 na Batería 2 (Táboa 6.11.), e entre 973 e 1217 palabras ós que habitualmente utilizan as dúas linguas (Táboa 6.12.). Na mostra de 6.º de Educación Primaria, estas diferenzas van desde 920 palabras na Batería 1 a 1414 na Batería 2, respecto dos que teñen como lingua preferente o castelán (Táboa 6.13.), mentres que superan ós que usualmente utilizan as dúas linguas entre 705 e 996 palabras nas Baterías 1 e 2, respectivamente (Táboa 6.14.). No que se refire á mostra de 2.º de ESO, o grupo galegofalante supera ó castelanfalante entre 1324 palabras na Batería 1 e 1568 na Batería 2 (Táboa 6.15.); a vantaxe en relación co grupo que adoita utiliza-las dúas está entre 878 na Batería 1 e 1246 palabras na Batería 2 (Táboa 6.16.).

Táboa 6.11.: Estatísticos descritivos da estimación do vocabulario na mostra total segundo a lingua familiar habitual (galego vs castelán)

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 galego	155	206	12.476	6546,15	2542,62	204,23
castelán	240	206	11.445	5299,48	2511,32	162,10
Batería 2 galego	155	206	14.538	7189,49	2733,33	219,55
castelán	240	206	12.640	5828,62	2850,81	184,02
Total Test galego	155	206	13.507	6867,32	2512,62	201,82
castelán	240	206	11.702	5563,86	2542,98	164,15

Táboa 6.12.: Estatísticos descritivos da estimación do vocabulario na mostra total segundo a lingua familiar habitual (galego vs as dúas linguas)

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 galego	155	206	12.476	6546,15	2542,62	204,23
	as dúas	180	1031	12.991	5573,76	2418,71
Batería 2 galego	155	206	14.538	7189,49	2733,33	219,55
	as dúas	180	618	13.507	5927,29	2627,89
Total Test galego	155	206	13.507	6867,32	2512,62	201,82
	as dúas	180	825	13.249	5750,42	2382,26

Táboa 6.13.: Estatísticos descritivos da estimación do vocabulario na mostra de 6.º de E. Primaria segundo a lingua familiar habitual (galego vs castelán)

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 galego	76	206	10.620	5955,68	2461,46	282,35
	castelán	116	206	10.826	5035,94	2312,14
Batería 2 galego	76	206	11.754	6669,88	2660,89	305,22
	castelán	116	206	12.373	5255,00	2627,04
Total Test galego	76	206	11.187	6311,88	2447,84	280,79
	castelán	116	206	10.620	5145,25	2325,23

Táboa 6.14.: Estatísticos descritivos da estimación do vocabulario na mostra de 6.º de Educación Primaria segundo a lingua familiar habitual (galego vs as dúas linguas)

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 galego	76	206	10.620	5955,68	2461,46	282,35
	as dúas	121	1031	10.620	5250,40	2341,48
Batería 2 galego	76	206	11.754	6669,88	2660,89	305,22
	as dúas	121	825	10.723	5675,75	2387,62
Total Test galego	76	206	11.187	6311,88	2447,84	280,79
	as dúas	121	928	10.569	5462,97	2229,22

Táboa 6.15.: Estatísticos descritivos da estimación do vocabulario na mostra de 2.º de ESO segundo a lingua familiar habitual (galego vs castelán)

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 galego	79	1034	12.476	7114,19	2504,36	281,76
	124	618	11.445	5546,02	2670,10	239,78
Batería 2 galego	79	1443	14.538	7689,37	2725,15	306,60
	124	618	12.640	6365,23	2956,12	265,47
Total Test galego	79	1391	13.507	7401,67	2472,52	278,18
	124	618	11.702	5955,47	2681,52	240,81

Táboa 6.16.: Estatísticos descritivos da estimación do vocabulario na mostra de 2.º de ESO segundo a lingua familiar habitual (galego vs as dúas linguas)

	N	Mínimo	Máximo	Media	Desv. típ.	Erro típ. da media
Batería 1 galego	79	1340	12.476	7114,19	2504,358	281,762
	59	1031	12.991	6236,93	2458,173	320,118
Batería 2 galego	79	1443	14.538	7689,37	2725,155	306,604
	59	618	13.507	6443,17	3019,144	393,059
Total Test galego	79	1391	13.507	7401,67	2472,519	278,180
	59	825	13.249	6339,93	2589,841	337,169

6.2. Baremos

Para que o test poida empregarse coa segunda finalidade (comparar suxeitos), preséntanse uns baremos con carácter provisional (Anexo IX). Estes baremos obtivéronse transformando o número de palabras estimadas coñecidas, entendidas como puntuacións directas, en puntuacións derivadas que permiten situar a un determinado suxeito no seu grupo, é dicir, indican a súa posición relativa nese grupo. Empregouse, con este fin, a escala derivada máis coñecida no contexto escolar: a centil, de tipo ordinal, que abarca de 1 a 99 e co termo medio en 50.

Os baremos, que precisan ser contrastados e perfeccionados con estudos posteriores, foron elaborados atendendo unicamente ó criterio diferenciador da idade. Non obstante, sería interesante que en ulteriores traballos os baremos se elaborasen tamén para outros niveis escolares e para outros criterios de diferenciación dentro de cada nivel educativo, como son o ámbito rural/urbano, costa/interior ou o medio socioeconómico de procedencia. En todo caso, a elaboración dos devanditos baremos podería levarase a cabo en función das variables con maior tamaño do efecto.

Como se pode comprobar no Anexo IX, os citados baremos refírense ás puntuacións das Batería 1 e 2, así como do Total do Test. Na parte inferior da Táboa encóntranse os estatísticos Número de suxeitos (N), Media e Desviación Típica das puntuacións. A consulta resulta sinxela, como en calquera outra táboa destas características. Nas columnas da esquerda e dereita sitúanse os centís, mentres que nas centrais están os valores correspondentes á estimación do número de palabras coas Baterías 1, 2 e co Total do Test. De tódalas formas, en tanto que probablemente sexa máis adecuado traballar cun menor grao de precisión, con rangos máis amplos, posiblemente sexa preferible utilizar estes baremos adaptándoos a unha escala como pode ser unha en quartís ou decís.

Para mostra-la secuencia completa a seguir, expónse a continuación un exemplo. Supoñamos que un docente ou un orientador está interesado en estudar en que medida aumenta o vocabulario dos escolares do seu centro e que factores puideran se-los responsables do aumento (volume de lecturas, estratexias didácticas en áreas como a de coñecemento do medio, intelixencia, etc.), co fin de programar unha intervención máis eficaz. Decide aplicar este test nun grupo de 6.º de ensinanza primaria e as estimacións medias do vocabulario son de 4227 palabras na Batería 1 (PC=35), 3712 na Batería 2 (PC=25) e 3969 palabras no total do Test (PC=30), que sitúan ó grupo por debaixo da media do seu nivel de escolarización. Cabe a posibilidade de intentar incidir en determinadas variables con obxecto de comprobar en que medida inflúen no aumento do vocabulario.

Comparando as estimacións de vocabulario que figuran no Anexo IX e no apartado 6.1 coas obtidas nalgúnhas das investigacións realizadas en lingua española, pódese comprobar que, para escolares con idades semellantes ós deste traballo, os datos non son coincidentes, se ben, como xa se expuxo no seu momento, en parte das mesmas tamén son diferentes tanto os tipos de mostras coma os procedementos metodolóxicos utilizados, razón pola que dificilmente se pode entender de forma estrita a comparación. Alén disto, convén ter en conta que hai investigacións que parecen mostrar que o vocabulario en castelán dos escolares galegos é máis amplo que o seu propio vocabulario en galego (Esteban, 1997) e que o seu vocabulario en castelán é máis reducido que o vocabulario dos escolares castelanfalantes de fóra de Galicia (Esteban, 1997; Suárez, Seisdedos e Meara, 1998).

Polo que respecta a esta investigación, a estimación do vocabulario coñecido como media para os escolares de sexto de Educación Primaria é de 5551 palabras e para o alumnado de segundo de ESO é de 6478 (Táboa 6.1.). Noutros estudos, como o de García Hoz (1946) realizado con alumnado da cidade de Madrid, se se calcula a media dos datos que achega, a estimación do vocabulario para os 11 anos é de 11.714 palabras e para os 13 anos é de 22.209. Non obstante, debe terse en consideración que neste último estudo a mostra estaba composta por escolares que cursaban o antigo bacharelato, polo que, a teor da selección que precedía á súa incorporación a esta etapa educativa, parece razoable supoñer que o nivel de coñecementos en xeral, incluído o vocabulario, seguramente fose superior ó dos escolares da mostra utilizada nesta tese, aínda que as idades coincidan. Pola súa parte, Justicia (1985a, 1985b) ofrece unha estimación do vocabulario activo de 4369 vocábulos para o grupo de escolares de 10 anos. Pero, neste caso, é preciso resaltar que non se trata dunha estimación do número de palabras que coñece un suxeito desta idade, senón do total de palabras usadas polo alumnado de 10 anos que compoñía a mostra empregada. Na investigación de Suárez, Seisdedos e Meara (1998), que si coincide metodoloxicamente coa investigación obxecto desta tese, a estimación para o alumnado de

11 anos, calculada a partir dos datos obtidos coas mostras de 9-10 anos e 13-14 anos, é de 7800 palabras, mentres que para o alumnado de 13 anos é de 9800 palabras.

A información que proporcionan os datos, tanto no referido ó volume de vocabulario coñecido coma no que se refire á situación relativa dun suxeito concreto, pode ser utilizada con diferentes finalidades. Por exemplo:

1. Como referencia, a fin de adapta-lo vocabulario ó nivel do alumnado, tanto na interacción verbal oral coma na selección de libros de texto. Parece razoable que nun centro no que a media da estimación do vocabulario do alumnado de 6.º de Educación Primaria, por exemplo, se situase nos percentís inferiores, sería conveniente seleccionar material didáctico que se axustase a este tamaño do vocabulario e que seguramente debería ser diferente ó doutro centro no que o alumnado se situase nos percentís superiores.

O estudo de Suárez, Seisdedos e Meara (1998) pode servir como mostra da posibilidade de detectar estas diferenzas no tamaño do vocabulario en castelán. Nesta investigación, a estimación do vocabulario dos escolares galegos sitúase entre os centís 44 e 46, mentres que o tamaño do vocabulario da mostra de Madrid sitúase entre os centís 60 e 63. Mesmo dentro de Galicia, os escolares pertencentes ós medios socioeconómicos medio-baixo e baixo sitúanse no centil 32, mentres que os escolares procedentes dos niveis medio-alto e alto sitúanse no 56. Igualmente, os mesmos autores encontraron o número de palabras coñecidas polo alumnado de centros urbanos era superior que o do alumnado de centros emprazados nunha zona rural de interior: a máis extrema foi de practicamente o dobre. Na investigación realizada para esta tese tamén se obtiveron diferenzas entre os escolares de diversos ámbitos: rural e urbano, costa e interior, etc. (punto 5.5.2).

2. Como dato para detectar escolares con diferentes graos de coñecemento do vocabulario, a fin de planear programas de enriquecemento e mellora, de forma individual ou grupal, se o nivel do mesmo o requirise.

3. Para contrastar a evolución de escolares sometidos a un programa de intervención para a mellora do vocabulario. Pero, o emprego con esta finalidade presenta unha limitación, como se indicou no punto 3.2.3 (capítulo 3), porque existe a posibilidade de que o suxeito sinala unha palabra sen coñecer realmente o seu significado se, para efectuar a estimación do número de palabras aprendidas, se utilizan as mesmas palabras que as usadas na instrución. Por outra parte, unha comprobación da evolución do vocabulario seguramente tería que realizarse tras unha intervención de longa duración.

Con todo, á hora de adoptar a decisión de considerar válida a aplicación e que os datos obtidos poden resultar de utilidade co fin de extraer conclusións de cara á intervención, é de suma importancia ter en conta o sinalado nos apartados 1 e 3 do capítulo 5, onde se explicaban as situacións nas que se decidiu descartar a un determinado número de escolares da mostra final.

Unha delas é a presenza de dificultades de lectura, xa que o test é dificilmente aplicable nas condicións habituais a alumnado que presente dificultades neste eido. Por iso, perante este tipo de situacións é necesario desestimar os datos ou, en todo caso, ampliar o espazo de tempo para a realización, se nos interesa aplicalo de todas as formas.

Outra das situacións está en relación cos casos en que non sexa posible obter unha estimación do vocabulario porque sinalan un excesivo número de pseudopalabras. En tales circunstancias, sempre que exista a posibilidade, é preciso repetir a aplicación e, de non ser posible, debe desestimarse o protocolo.

Unha terceira situación refírese ós casos en que a estimación do vocabulario nunha das baterías resulta ser dobre ou máis que na outra. Cando se produza esta situación, será necesario desestimar os protocolos ou ben repetir a aplicación, se é factible.

Finalmente, cabe incidir en que, como se suxeriu con anterioridade, para obter unha estimación máis fiable do tamaño do vocabulario, en principio, se só se emprega unha

Batería, é conveniente indicar con cá se acadou a estimación e, sempre que existe a posibilidade, é preferible efectua-la estimación utilizando o test completo.

Capítulo 7

Conclusións

Non existían en Galicia investigacións, cando menos que souberamos, que abordaran a estimación do volume do vocabulario galego coñecido por escolares galegos, nin probas específicas que puideran ser utilizadas con esta finalidade.

Ante esta situación, neste traballo propuxémonos construír un instrumento o suficientemente fiable e válido que permitira efectuar unha estimación do tamaño do vocabulario coñecido en lingua galega polo alumnado de Galicia entre 8 e 16 anos. Entre as diferentes opcións, en canto a procedementos e probas xa existentes noutros contextos, inclinámonos pola utilización dunha proba de resposta tipo Si/Non, en tanto que permite inferir-lo tamaño do vocabulario e que, se ben estas probas presentan puntos débiles (considérase que miden esencial ou soamente coñecemento pasivo, non permiten diferenciar distintos significados dunha palabra, a elección de estímulos falsos pode afectar seriamente ás puntuacións, presentan limitacións para avaliar vocabulario despois dun período de instrución, xa que poden darse respostas máis por familiaridade que por coñecemento do significado ...), presentan tamén vantaxes (son fáciles de construír, as instrucións son fáciles de entender, requiren pouco tempo para a súa aplicación, permiten probar un número considerable de ítems, co cal pódense facer estimacións máis fiables do número total de palabras coñecidas, simplifican a tarefa que se require para comprobar se se coñece unha palabra, son susceptibles de corrección mecánica, etc.).

Para a súa elaboración seguíronse de cerca as contribucións de Meara e Buxton (1987), Goulden Nation e Read (1993) e de Suárez, Seisdedos e Meara (1998), aínda que tendo en conta unha ampla contextualización (Parte primeira da tese). En primeiro lugar,

seleccionouse ó chou unha mostra de palabras do dicionario da Real Academia Galega (1997). En segundo lugar, mediante a valoración de tres xuíces, clasificáronse as palabras en catro categorías, atendendo á súa suposta dificultade. En terceiro lugar, determinouse a proporción de palabras de cada categoría que formarían parte de cada un dos 10 subtests que se decidiu construír. En cuarto lugar, xeráronse pseudopalabras, que foron valoradas por dous xuíces, co fin de eliminar aquelas que non terían cabida no idioma galego. En quinto lugar, seleccionáronse 20 pseudopalabras ó azar para cada subtest. Finalmente, mesturáronse aleatoriamente os 60 ítems asignados a cada subtest.

Optouse por traballar con alumnado de 6.º de Educación Primaria e de 2.º de ESO, en tanto que son dous dos momentos cruciais no contexto da escolaridade obrigatoria. Para comprobar se eran comprensibles e de utilidade as instrucións do test, así como se era manexable o formato do test, comezouse o proceso con dúas aplicacións piloto a alumnado de 6.º de Educación Primaria e de 2.º de ESO. En base ós datos obtidos, leváronse a cabo axustes tanto nas instrucións coma no formato das probas.

Despois dunha análise do grao de coñecemento da lingua galega por parte da poboación, utilizando os datos do Mapa Sociolingüístico de Galicia e doutras investigacións, coa finalidade de verifica-la súa utilidade en canto á capacidade para discriminar entre diferentes grupos, formuláronse cinco hipóteses no sentido de que o test sería sensible ás diferenzas no tamaño do vocabulario entre alumnado pertencente a diferentes grupos de poboación.

Para a administración definitiva, seleccionáronse centros con alumnado de 6.º de Educación Primaria e 2.º de ESO de titularidade pública e privada, pertencentes a zonas do interior e da costa, así como de ámbito rural e urbano.

As dificultades expostas no seu momento no relativo á configuración da mostra, que impediron a equiparación numérica entre os diferentes agrupamentos, pero tamén no que se

refire á aplicación das probas ou no que atinxe ós descartes de suxeitos que foi necesario efectuar, limitan as posibilidades de chegar a conclusións máis amplas e fundamentadas en base ós datos obtidos. Con todo, estes datos permiten ofrecer algunhas conclusións.

Conseguíuse a finalidade que se propoñía para o traballo: construír un instrumento co que de forma fiable e válida fose posible estima-lo número de palabras que coñecen en lingua galega, ao menos, os escolares galegos de 6.º de Educación Primaria e de 2.º de ESO.

Das análises realizadas neste traballo, despréndese o seguinte:

a) A fiabilidade do Test Total e das Baterías por separado, tendo en conta as correlacións obtidas, resulta aceptable.

b) Polo que respecta á validez, a de construto pódese considerar que queda avalada pola fundamentación teórica, así como polas consideracións expostas en canto á descrición da mostra. Esta validez queda igualmente avalada polos resultados da comprobación das cinco hipóteses. A teor destes resultados, é significativamente maior o tamaño do vocabulario dos escolares de 2.º de ESO que o dos escolares de 6.º de Educación Primaria; o vocabulario do alumnado rural e de interior é maior que o vocabulario do alumnado urbano e de costa, respectivamente; o vocabulario dos suxeitos escolarizados en centros públicos é maior que o dos que están escolarizados en centros privados; finalmente, é maior o tamaño do vocabulario do alumnado que ten o galego como lingua familiar que o do alumnado que ten o castelán como lingua familiar ou que o do alumnado que fala as dúas linguas.

A validez de contido pode considerarse que queda avalada polo proceso de construción do test. Ademais, o grao de dificultade das palabras está en consonancia coa valoración que da mesma efectuaron profesores de ensino non universitario.

A validez de criterio ou concorrente, calculada correlacionando o vocabulario estimado co rendemento global na área de lingua galega, consonte á apreciación do

profesorado da materia, alcanza uns valores que se aproximan ou superan unha correlación do 0,50, utilizando as estimacións a partir das puntuacións nas Baterías e no total do Test. Se se utilizan as correlacións dos subtests, os valores son inferiores, pero significativas, situándose en termos xerais a partir de 0,30.

Non obstante, o test preséntase como unha proposta experimental, que debería ser contrastada con diferentes mostras en posteriores investigacións desta índole, acumulando evidencia sobre o seu funcionamento na práctica, antes de outorgarlle un carácter máis definitivo. Poderíase ter realizado máis elaboracións dos datos, pero considerouse suficiente para unha primeira versión. De tódolos modos, aínda con esta etiqueta de “experimental”, pode constituír unha contribución de certa importancia por diferentes razóns:

Como se expuxo no capítulo relativo á súa relevancia, a riqueza do vocabulario dunha persoa pode considerarse un indicador da súa riqueza de coñecementos do mundo o que constitúe unha información de gran importancia pedagóxica.

O vocabulario, alén da súa interrelación co desenvolvemento cognitivo en xeral, constitúe un elemento básico da linguaxe verbal, tanto no plano oral coma no escrito, tanto no aspecto receptivo coma no produtivo, e a linguaxe verbal é a ferramenta curricular fundamental, xogando un papel esencial no rendemento escolar.

Sirva a modo de exemplo a interrelación que existe co manexo da linguaxe escrita, esencialmente coa comprensión lectora. Neste sentido, malia que as variables que determinan o grao de comprensión dun texto escrito son múltiples (memoria de traballo, coñecemento sintáctico, coñecemento xeral do mundo e sobre a estrutura dos textos, habilidade para activa-los coñecementos previos, así como o coñecemento e o vocabulario sobre o tema en cuestión, etc.), a importancia do vocabulario deriva de que a súa riqueza reflicte parte do coñecemento do mundo de que se dispón, e a activación deste coñecemento relacionado co contido do texto que se le resulta fundamental para a comprensión lectora. Deste modo, o coñecemento do vocabulario parece representa-lo

primeiro chanzo no complexo proceso de comprensión dun texto, xa que se descoñecen as palabras, e máis se son elementos fundamentais, dificilmente se poderá progresar desde o nivel de comprensión superficial a un nivel de comprensión máis profundo. Ó respecto, coñecer unha estimación do tamaño do vocabulario pode servir de axuda para o labor docente, en tanto que, cando menos, un primeiro indicador para tratar de acomoda-los materiais curriculares ás posibilidades do alumnado en canto a vocabulario.

Aínda que o test elaborado é fundamentalmente unha replicación doutros xa existentes, polo que non representa unha novidade en si mesmo, si é o primeiro traballo levado a cabo co propósito de estima-lo tamaño do vocabulario en galego de alumnado escolarizado en Galicia, xa que outros como é o caso do *Threshold Level Galego Vocabulary Test* está deseñado para ser utilizado con estudantes de galego como lingua estranxeira. É de esperar que o estudo efectuado nesta tese e o test construído contribúan á realización doutras investigacións sobre o coñecemento do vocabulario en lingua galega.

No proceso de análise da validez de constructo obtivéronse unha serie de resultados que poden resultar de interese comentar:

Hai diferenzas significativas no tamaño do vocabulario en función da idade, tal como se formulou na primeira hipótese. Confirmouse, en consecuencia, que é significativamente maior o tamaño do vocabulario do alumnado de máis idade e que está escolarizado en dous niveis académicos superiores. Tendo en conta esto, pódese pensar que o vocabulario dos escolares galegos aumenta nunhas 1000 palabras de media aproximadamente entre 6.º de Educación Primaria e 2.º de ESO (Táboa 6.1.).

De tódolos modos, existe unha considerable variabilidade entre escolares dunha mesma idade e por esta razón se elaboraron uns baremos elementais en puntuacións centís.

Tamén se indicou xa anteriormente que hai diferenzas significativas na mostra total entre o grupo de poboación rural e o de urbana, entre o de costa e o de interior, entre o

alumnado de centros públicos e privados e segundo a lingua familiar habitual. Confirmouse ó respecto que é maior o tamaño do vocabulario do alumnado rural que o do urbano, o do alumnado de interior que o do de costa, o dos escolares de centros públicos que o dos de centros privados e o dos escolares co galego como lingua habitual.

Analizando por separado os datos da mostra de alumnado de 2.º de ESO en base ós trazados dos gráficos correspondentes (Figuras 5.13. a 5.24.), pódese pensar que se produce unha tendencia á existencia de diferenzas entre os escolares que pertencen á zona rural e á urbana, entre os que pertencen á zona de costa e á de interior, entre os sectores de poboación representados pola titularidade dos centros, así como en función da lingua utilizada de forma preferente no ámbito familiar.

No relativo á mostra de 6.º de Educación Primaria, os trazados gráficos unicamente permiten pensar que poidan producirse diferenzas claras segundo a lingua habitual. Polo contrario, os datos representados nos gráficos non indican que poidan existir diferenzas claras entre os grupos rural/urbano, costeiro/interior e segundo a titularidade dos centros.

A explicación da tendencia a que se produzan diferenzas máis claras na mostra de 2.º de ESO, as cales se reducen considerablemente en 6.º de Educación Primaria, cabe a posibilidade de atribuíla a distintas causas, entre as que se poderían sinalar:

a) Que sexa unha reflexo do que ocorre a nivel macrosocial (proceso de urbanización e desgaleguización, consistente nunha perda gradual do galego no ámbito familiar, efecto máis notorio canto máis novos son os individuos, xunto con canto máis alta é a clase social e maior é o carácter urbano do lugar de nacemento). O sistema escolar convértese na canle alternativa de transmisión do galego para este sector poboacional (Fernández e Rodríguez, 1995). Neste sentido, canto máis se produce a circunstancia de que o galego máis habitual sexa o escolar, é probable que se produza progresivamente un efecto de igualación no grao do seu coñecemento, e, entre este, do vocabulario.

b) Outra posible explicación, podería ser que se produza un cambio nos individuos coa idade en canto ó coñecemento do galego, tendo en conta que traballamos con vocabulario escrito, de forma que entre os máis novos en idade as diferenzas sexan de menor importancia e, co aumento da idade, estas diferenzas se acrecenten substancialmente. Sucedería algo semellante ó que ocorre coas diferenzas en canto a coñecementos académicos en xeral, que semella elevarse coa idade.

Finalmente, cabe sinalar algunhas suxestións acerca de como utiliza-lo instrumento na práctica e como sería posible ir acumulando evidencias que permitan chegar a un pronunciamento máis definitivo sobre a súa validez e utilidade.

Optimiza-las condicións de administración no relativo á uniformación do procedemento, co que se podería conseguir incrementa-lo número de protocolos de reposta válidos, é dicir, nos que a puntuación non resultase ser inferior a 1. Neste sentido, é sumamente importante asegurar que os suxeitos sinalen os ítems só cando estean realmente seguros de coñecer unha palabra e non se produza a conduta de sinalar un número elevado sen esta seguridade, coa crenza de que sinalando un número maior terán máis posibilidades de "sacar maior puntuación". Ó respecto, detectouse o enorme peso das instrucións nesta investigación e cabe resaltar que resulta efectivo, por exemplo, ir recordándolles cada certo espazo de tempo a suxestión de ser moi coidadosos neste sentido.

Reiterar que probablemente conveña aplica-las dúas Baterías (considera-las dúas como un único test), o que repercute nunha maior fiabilidade e estabilidade das estimacións.

Mellora-las mostras, posto que como xa se sinalou, debido tanto ás posibilidades de selección como ás dificultades xurdidas polas condicións en que se aplicaron, quedaron desequilibradas en canto ó número de alumnado que compoñía os agrupamentos establecidos.

Levar a cabo investigacións con diferentes estratos de idade que permitan ir efectuando os axustes necesarios no que se refire á elaboración do test. Algúns exemplos deste tipo de investigacións, encadradas tanto no contexto do paradigma da Investigación-Acción por parte de profesionais de orientación ou do profesorado, coma desde outros, son utiliza-la súa aplicación para estudar:

- A relación co rendemento escolar xeral e/ou por áreas, así como co inicio da aprendizaxe da lingua escrita, coa comprensión lectora ou coa expresión escrita.
- A relación entre a riqueza de vocabulario e determinadas características socioeconómicas, culturais, psicolóxicas ou de xénero.
- O seu incremento ó longo dun período evolutivo concreto.
- A efectividade de programas de intervención, individuais ou grupais.

Podería ser igualmente interesante elaborar tests de tamaño do vocabulario específico para determinados ámbitos de coñecemento.

Para rematar, cabe sinalar que se pode considerar que, no fundamental, conseguíuse alcanza-lo propósito xeral desta investigación, cos matices e limitacións que se foron expoñendo ó longo do traballo. Con elo e coa profundización no ámbito do coñecemento do vocabulario e a súa valoración que esixe un traballo destas características, esta investigación tamén representa persoalmente un avance nunha temática á que dun modo ou outro estou vinculado profesionalmente desde o inicio da miña carreira profesional.

Bibliografía

-
- Adams, A. K. (1986). Sentence frame effects on children's category judgments. *Papers and Reports on Children Language Development*, 25, 1-8.
- Aguado, G. (1995). *El desarrollo del lenguaje de 0 a 3 años. Bases para un diseño curricular en Educación infantil*. Madrid: CEPE.
- Aguado, G. (1999). *Trastorno específico del lenguaje. Retraso de lenguaje y disfasia*. Málaga: Aljibe.
- Aitchison, J. (1987). *Words in the mind: an introduction to the mental lexicon*. Oxford: Basil Blackwell.
- Alabama State Department of Education (2001). *Oral Language and Vocabulary Development: Grades 2-3. Alabama Reading Initiative*. Montgomery, AL: Alabama Reading Initiative (ERIC ED 464.374).
- Alameda J. R. e Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Universidad de Oviedo.
- Alonso M., A. (1996). *Glosario de la terminología gramatical*. Madrid: Magisterio Español.
- Alonso, J., Carriedo, N. e González, E. (1992). Evaluación de la capacidad de comprender y resumir lo importante: la batería IDEPA. En J. Alonso (Dir.), *Leer, comprender y pensar. Nuevas estrategias y técnicas de evaluación (pp. 59-103)*. Madrid: MEC/CIDE.
- Alonso T., J. e Mateos, M^a. M. (1985). Comprensión lectora: modelos, entrenamiento, evolución. *Infancia y Aprendizaje*, 31-32, 5-30.
- Alpeter, T. S. (1985). Use of PPVT-R for intellectual screening with a preschool pediatric sample. *Journal of Pediatric Psychology*, 10 (2), 195-198.
- Alpeter, T. S. (1989). The PPVT-R as measure of psycholinguistic functioning: A caution. *Journal of Clinical Psychology*, 45 (6), 935-941.
- Alpeter, T. e Handal, P. J. (1986). Use of the PPVT-R for intellectual screening with school-age children: A caution. *Journal of Psychoeducational Assessment*, 4 (2), 145-154.
- Alpeter, T. S. e Johnson, K. A. (1989). Use of the PPVT-R for intellectual screening with adults: A caution. *Journal of Psychoeducational Assessment*, 7 (1), 39-45.
- Alvar E., M. (1993). *Lexicografía descriptiva*. Barcelona: Bibliograf.
- Álvarez, C. J., Alameda, R. e Domínguez, A. (1999). El reconocimiento de las palabras: el procesamiento ortográfico y silábico. En M. de Vega e F. Cuetos (Coords.), *Psicolingüística del español (pp. 89-130)*. Madrid: Trotta.
- Álvarez, C. e Díez-Itza, E. (2000). Competencia léxica y rendimiento académico en alumnos de segundo de bachillerato. *Aula Abierta*, 7, 185-195.
- Anastasi, A. (1977). *Tests psicológicos*. Madrid: Aguilar.
-

- Anderson, R. C. e Freebody, P. (1979). *Vocabulary Knowledge*. Urbana, IL: University Illinois, Center for the Study of Reading. (Inf. Tecn. Nº 136). (ERIC ED 177.480).
- Anderson, R. C. e Freebody, P. (1983). Reading comprehension and the assesment and acquisition of word knowledge. En B. Hutson (Ed.), *Advances in Reading Language Research: A research annual*. Vol. 2, (pp. 231-256). Greenwich, Ct: JAI Press.
- Anderson, R. C. e Nagy, W. E. (1991). Words meanings. En R. Barr, M.L. Kamil, P.B. Mosenthal e P.D. Pearson (Eds.), *Handbook of Reading Research*, Vol. II (pp. 620-724). New York: Longman.
- Anderson, R. C. e Nagy, W. E. (1993). *The Vocabulary Conundrum*. Urbana, IL: Center for the Study of Reading. (Inf. Técn. Nº 570). (ERIC ED 354.489).
- Anglin, J. M. (1970). *The Growth of Word Meaning*. Cambridge, Mass.: MIT Press.
- Anglin, J. M. (1977). *Word, object and conceptual development*. New York: Norton.
- Anglin, J. M. (1993). Vocabulary development. A morphological analysis. *Monographs of the Society for Research in Child Development*, Serial nº 238, Vol. 58, nº 10.
- Armayer, H. (1975). *Determinación del vocabulario fundamental (de uso y de reconocimiento) del niño cordobés de seis años*. Córdoba: Servicio de Publicaciones del I.C.E. de la Universidad.
- Armayer, H. (1977). *Determinación del vocabulario fundamental (de uso y de reconocimiento) del niño cordobés de 14 años, en el último curso de E.G.B. y su incorporación al B.U.P.* Córdoba: Servicio de Publicaciones del I.C.E. de la Universidad.
- Arza, N., Campos, A., Veiga, D. e Rubal, X. (1990). A situación do proceso de normalización lingüística no ensino. En J. A. Caride (Dir.), *A educación en Galicia. Informe Cero. Problemas e Perspectivas* (pp. 97-142). Santiago de Compostela: ICE Universidade.
- Arza, N., Rubal, X. e Veiga, D. (1992). *Aproximación á situación da lingua no ensino non universitario: preescolar e EXB*. Santiago de Compostela: Consellería de Educación e Ordenación Universitaria, Xunta de Galicia.
- Asaka, T. (1996). *Vocabulario básico da lingua galega*. Tokio: Daigakusyorin.
- August, D. e Hakuta, K. (Eds.) (1997). *Improoving Schooling for Language-Minority Children: A Research Agenda*. National Research Council, Institute of Medicine, National Academic Press.
- Averril, L. A. (1956). *La vida psíquica del escolar*. Buenos Aires: Paidós.
- Azorín, F. (1994). *Métodos y aplicaciones del muestreo*. Madrid: Alianza Universidad Textos.
- Baddeley, A., Logie, R. e Nimmo, S. I. (1985). Components of fluent reading. *Journal of Memory and Language*, 24 (1), 119-131.

-
- Bauer, L. e Nation, P. (1993). Word Families. *International Journal Of Lexicography*, 6 (4), 253-279.
- Beck, J. (1981). *New vocabulary and the associations it provokes*. Proxecto presentado para a obtención do M.A. in Second Language Learning and Teaching. Birkbeck College, London. Citado en Suárez (1987).
- Beck, F. W., Black, F. L. e Doles, J. (1985). The concurrent validity of the PPVT-R relative to the Comprehensive Test of Basic Skills. *Educational and Psychological Measurement*, 45 (3), 705-710.
- Beck, I. e McKeown, M. G. (1991). Conditions of vocabulary acquisition. En R. Barr, M.L. Kamil, P. Mosenthal e P.D. Pearson (Eds.), *Handbook of Reading Research. Vol. I (pp. 789-814)*. New York: Longman.
- Belinchón, M., Rivière, A. e Igoa, J. M. (1992). *Psicología del lenguaje. Investigación y teoría*. Madrid: Trotta.
- Benedict, H. (1979). Early lexical development comprehension and production. *Journal of Child Language*, 6, 183-200.
- Bennet, G. M., Bennet, M. J., Clendenen, D.M., Doppelt, J. E., Ricks, J. H., Seashore, H. G. et al. (1985). *Test de Pronóstico Académico*. APT. Madrid: TEA.
- Blanche-Benveniste, C. (1998). *Estudios lingüísticos entre oralidad y escritura*. Barcelona: Gedisa.
- Bloom, P. (2002). *How children learn the meaning of words*. Cambridge, MA: MIT Press.
- Bloomfield, L. (1933). *Language*. New Cork: Holt (Edición en español, *Lenguaje*. Lima, Perú: Universidad Nacional Mayor de San Marcos, 1964).
- Boehm, A. E. (1988). *Test Boehm de Conceptos Básicos*. Madrid: TEA.
- Brandín, C. (2002). Velocidade de procesamento da información de material escrito en bilingües galegos. En M^a. X. Bugarín, J. Cajide, A. Dosil, X. G. Ferreiro, M. González e M. A. Santos (Eds.), *Actas da VIII Conferencia Internacional de Linguas Minoritarias* (pp. 405-412). 22-24 novembro de 2001. Santiago de Compostela: Xunta de Galicia.
- Bruner, J. S. (1978). On prelinguistic prerequisites of speech. En R. N. Campbell e P. T. Smith (Eds), *Recent Advances in the Psychology of Language Development and Mother-Child Interaction (pp. 199-214)*. New York: Plenum Press.
- Bruner, J. S. (1986). *El habla del niño*. Barcelona: Paidós.
- Bruning, R. H., Schraw, G. J., Norby, M. N. e Rowning, R. R. (2005). *Psicología cognitiva y de la instrucción*. Madrid: Pearson-Prentice Hall.
- Brynildssen, S. (2000). *Vocabulary's Influence in Successful Writing*. Bloomington, IN: ERIC Clearinghouse on Reading English and Colmmunication (ERIC ED 446.339).
-

- Buchanan, M. A. (1927). *A Graded Spanish Word Book*. Toronto.
- Butcher, H. J. (1973). *Sampling in Educational Research*. Manchester: Univetsity Press.
- Cabrera, F. A. e Espín, J. V. (1986). *Medición y evaluación educativa*. Barcelona: PPU.
- Cajide, X. (Dir.) (1989). *Modelos de ensino bilingüe axeitados á realidade galega*. Santiago de Compostela: Xunta de Galicia. Consellería de Educ. e Ordenación Universitaria.
- Callahan, D. e Drum, P. A. (1984). Reading ability and prior knowledge as predictors of eleven and twelve year olds' text comprehension. *Reading Psychology*, 5 (1-2), 145-154.
- Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Drestler, Ch., Lipmann, D. et al. (2004). Closing the gap: Addressing vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39 (2), 188-215.
- Carroll, L. (1973). *Alicia a través del espejo y lo que Alicia encontró al otro lado*. Madrid: Alianza Editorial.
- Carroll, J. B., Davies, P. e Richman, B. (1971). *The American heritage word frequency book*. Boston: Houghton Mifflin.
- Carter, R. (1986). Good Word! Vocabulary Style and Coherence in Children's Writing. En J. Harris e J. Wilkinson. *Reading Children Writing. A linguistic View (pp. 92-211)*. London: Allen e Unwin Publishers Ltd.
- Carvajal, H., Hardy, K., Harmon, K., Sellers, T. A (1987). Relationships among scores on the Stanford-Binet IV, Peabody Picture Vocabulary Test-Revised and Columbia Mental Maturity Scale. *Bulletin of the Psychonomic Society*, 25 (4), 275-276.
- Carvajal, H., McVey, S., Sellers, T. A e Weyand, K. (1987). Relationships among scores on the General Purpose Abbreviated Battery of Stanford-Binet IV, Peabody Picture Vocabulary Test-Revised, Columbia Mental Maturity Scale and Goodenough-Harris Drawing Test. *Psychological Record*, 37 (1), 127-130.
- Carvajal, H., Shaffer, C. e Weaver, K. A. (1989). Correlations of scores of maximun security inmates on Weschsler Adult Intelligence Scale-Revised and Peabody Picture Vocabulary Test-Revised. *Psychological Reports*, 65 (1), 268-270.
- Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behaviour*, 26, 413-437.
- Casanova, M^a. A. e Rivera, M. (1989). *Vocabulario básico en la EGB, I-III*. Madrid: MEC-Espasa Calpe.
- Cattell, R. B. e Cattell, K. S. (1989). *Test de Factor G*. Madrid: TEA.
- Centro Ramón Piñeiro para a Investigación e Humanidades: *Corpus de Referencia do Galego Actual (CORGA)*. <http://corpus.cirp.es/corgaxml>.

-
- Chall, J. S. (1987). Two vocabularies of reading: Recognition and meaning. En M.G. McKeown e M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 7-17). Hillsdale, NJ: Erlbaum.
- Chamberlain, A. (1965). Learning a passive vocabulary. En *International Conference Modern Foreign Language Teaching*, Noviembre, Parte I. Citado en A. Suárez (1987).
- Chanell, J. (1988). Psycholinguistics considerations on the study of L2 vocabulary acquisition. En R. Carter e M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 83-96). London: Longman.
- Chao, Yuen Ren (1975). *Iniciación a la lingüística*. Madrid: Cátedra.
- Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, 15, 317-335.
- Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, 17, 417-431.
- Clark, E. V. (1995). *The lexicon in acquisition*. Cambridge, UK: Cambridge University Press.
- Clark, E. V. e Hecht, B. F. (1986). Comprehension, production and language acquisition. *Annual Review of Psychology*, 34, 325-349.
- Clark, H. H. (1975). Asociaciones de palabras y teoría lingüística. En J. Lyons (Ed.), *Nuevos horizontes de la lingüística* (pp. 285-300). Madrid: Alianza.
- Cohen, L. e Manion, L. (1990). *Métodos de investigación educativa*. Madrid: La Muralla.
- Cohen, L., Manion, L. e Morrison, K. (2001). *Research Methods in Education*. London: Routledge Falmer.
- Collins, A. M. e Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 82, 407-428.
- Coordinadora d'Ensenyament I Llengua Catalana (1978). *POAL. Vocabulari bàsic ampliat*. Barcelona: Ediciones Roca.
- Cooper, J. D. (1997). *Literacy. Helping Children Construct Meaning*. Boston: Houghton Mifflin Company.
- Corson, D. (1995). *Using English Words*. New York: Kluwer.
- Corson, D. (1997). The learning and use of academic English words. *Language Learning*, 47 (4), 671-718.
- Council of Europe (1986). *Nivel Llindar*. Barcelona: Generalitat de Catalunya. Política Lingüística.
- Cruz, M^a. V. de la (1982): *Batería de Aptitudes para el Aprendizaje Escolar*. Madrid: TEA.

- Crystal, D. (1991). Teaching Vocabulary: The Case for Semantic Curriculum. En K. Mogford-Bevan e J. Sadler (Eds.), *Child Language Disability. Vol. II: Semantic and Pragmatic Difficulties* (pp. 43-58). Clevedon, Avon (UK): Multilingual Matters Ltd.
- Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge, UK: Cambridge University Press.
- Cuetos, F. (1991). *Psicología de la escritura*. Madrid: Escuela Española.
- Cummins, J. (2002). *Lenguaje, poder y pedagogía*. Madrid: Morata.
- Dale, E. (1965). Vocabulary measurement: techniques and major finding. *Elementary English*, 36, 895-901.
- Dale, Ph. S. (1976). *Language Development. Structure and Function*. New York: Holt, Rinehart and Winston.
- Dale, Ph. S. (1991). The validity of the parent measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Research*, 34, 565-571.
- D'Anna, C. A. e Zechmeister, E. B. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behaviour*, 23, 1, 109-122.
- Diack, H. (1975). *Test Your Own Wordpower*. St. Albans: Paladin.
- Díaz, J. M. e De Vega, M. (2003). Modelos de situación e inferencias en la comprensión de textos. En J.A. León (Coord.) *Conocimiento y discurso. Claves para inferir y comprender* (pp. 139-152). Madrid: Pirámide.
- Díaz Castañón, C. (Dir.) (1977). *Vocabulario básico del español y sus aplicaciones a la enseñanza* (3 vols.). Oviedo: ICE Universidad.
- Diller, K. C. (1978). *The Language Teaching Controversy*. Rowley, Mass.: Newbury House.
- Dolby, J. L. e Renikoff, H. L. (1967): *The English Word Speculum*. The Hague: Mouton and Co.
- Dolch, E. W. (1936). How much word knowledge bring to grade 1?. *Elementary English Review*, 13, 177-183. Citado en D. D. Jonson (2000).
- Doman, G. (1991). *Como enseñar a leer a su bebé*. México: Diana.
- Domínguez, A. e Estévez, A. (1999). El reconocimiento de las palabras: procesamiento morfológico y semántico. En M. de Vega e F. Cuetos (Coords.), *Psicolingüística del español* (pp. 131-161). Madrid: Trotta.
- Downie, N. M. e Heath, R. W. (1983). *Métodos estadísticos aplicados*. Madrid: Ediciones del Castillo.

-
- Duin, A. H. (1983). *The effects of Intensive Vocabulary Instruction on a Specific Writing Task. A Plan B Paper* (ERIC ED 239.222).
- Duin, A. H. e Graves, M. F. (1986). Effects of vocabulary instruction used as prewriting technique. *Journal of Research and Development in Education*, 20, 7-13.
- Duin, A. H. e Graves, M. F. (1987). Intensive vocabulary instruction as prewriting technique. *Reading Research Quarterly*, 22 (3), 311-330.
- Duin, A. H. e Graves, M. F. (1988). Teaching vocabulary as a writing prompt. *Journal of Reading*, 32, 204-212.
- Dunn, L. M. (1985). *Test de vocabulario en imágenes Peabody*. Madrid: Mepsa.
- Dupuy, H. J. (1974). *The Rationale Development and Standardization of a Basic Word Vocabulary Test*. Washington, DC: US Government Printing Office. Citado en R. Goulden, P. Nation e J. Read (1990).
- Eco, U. (1994). *Six Walks in the Fictional Wood*. Cambridge: Harvard University Press.
- Eldredge, J. L. (1990). Causal Relationships between Phonics, Reading Comprehension and Vocabulary Achievement in the Second Grade. *Journal of Educational Research*, 83 (4), 201-214.
- Ellis, N. (1995). Vocabulary Acquisition: Psychological Perspectives. *Language Teachers*, 19 (2), 12-16.
- English, F. (1985). *Measuring vocabulary in non-native speakers of English*. Unpublished dissertation. London University (Birkbeck College).
- Escoriza, J. (1986). *Predicción del rendimiento en lectura en el primer curso de EGB*. Tese de doutoramento. Universidade Autónoma de Barcelona.
- Espín, J. V. (1986). Estudio predictivo del rendimiento en comprensión lectora de niños que finalizan el ciclo inicial en función de la clase social. *Revista de Innovación Educativa*, 7 (4), 21-34.
- Espín, J. V. (1987). *Lectura, lenguaje y educación compensatoria*. Barcelona: Oikos-Tau.
- Espin, Ch. e Foegen, A. (1996). Validity of General Outcome Measures for Predicting Secondary Student's Performance on Content-Areas Tasks. *Exceptional Children*, 62 (6), 497-514.
- Esteban R., S. (1997). *O vocabulario galego/castelán no ensino en Galicia. Unha aproximación socio-lingüística*. Ourense: Galiza Editora.
- Farley, M. J. e Elmore, P. B. (1992). The Relationship of Reading Comprehension to critical Thinking Skills, Cognitive Ability and Vocabulary for a Sample of Underachieving College. *Educational and Psychological Measurement*, 52 (4), 921-931.
-

- Fenson, L., Pale, P. S., Reznik, J. S., Thal, D., Bates, E., Hartung, J. et al. (1993). *The McArthur Communicative Development Inventories: user's guide and technical manual*. San Diego: Thomson Learning.
- Ferguson, G. A. (1971). *Statistical Analysis in Psychology and Education*. London: McGraw-Hill Kogakusha.
- Fernández, M. (1983). La situación de la lengua en Galicia: la lengua de los escolares. En M. Siguán (Coord.), *Lenguas y educación en el ámbito del estado español* (pp. 198-235). Barcelona: ICE Universidad de Barcelona.
- Fernández Ballesteros, R. e Carrobles, J. A. I. (1983). *Evaluación conductual*. Madrid: Pirámide.
- Fernández R., M. A. e Rodríguez N., M. A. (Coords.) (1994). *Lingua inicial e competencia lingüística en Galicia*. A Coruña: Seminario de Sociolingüística-RAG.
- Fernández R., M. A. e Rodríguez N., M. A. (Coords.) (1995). *Usos lingüísticos en Galicia*. A Coruña: Seminario de Sociolingüística-RAG.
- Fernández R., M. A. e Rodríguez N., M. A. (Coords.) (2000). Datos facilitados persoalmente en soporte informático sobre Lengua inicial, competencia lingüística e usos lingüísticos en Galicia.
- Fernández, X. A., Monteagudo, H. e Puga, M. (1993). *Nivel Soleira (Versión provisional)*. Santiago de Compostela: Xunta de Galicia. Consellería de Educación e Ordenación Universitaria.
- Ferraces, M^a. J. e Andrade, E. M. (s.d.). *Deseños de investigación. Esquemas*. Santiago de Compostela: Tórculo Edicións.
- Fletcher, P. (1987). Aspects of language development in the preschool years. En W. Yule e M. Rutter (Eds.) *Language Development and Disorders*. London: McKeith Press.
- Flower, L. e Hayes, J. R. (1994). A cognitive process theory of writing. En R.B. Ruddell, M.R. Ruddell e H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 928-950). Newark, DE: International Reading Association.
- Freebody, P. e Anderson, R. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly*, 18, 277-294.
- Fries, Ch. C. e Traver, A. A. (1960). *English Word List*. Ann Arbor, Michigan: George Wahr Publishing.
- Fruyt, M. e Reichler-Beguelin, M. J. (1990). La notion de "mot" en latin et dans d'autres langues indoeuropéennes anciennes. *Modèles linguistiques*, 12 (1), 21-46.
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M. e Wenger, N.R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 28, 84-89.

-
- Golinkoff, R. M., Mervis, C. B. e Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, 21, 125-155.
- Galisson, R. e Coste, D. (1976). *Dictionnaire de didactique des langues*. Paris: Hachette.
- García R., D. e Meara, P. (1996). *Threshold Level Galego Vocabulary Tests*. Swansea: Centre for Applied Language Studies University of Wales.
- García Hoz, V. (1946). Evolución cuantitativa del vocabulario en escolares de 9 a 18 años. *Revista Española de Pedagogía*, 13-14, 403-434.
- García Hoz, V. (1952). El vocabulario común y su utilización didáctica. *Bordón*. 4, 29. (Reeditado en 1999 en Bordón, 51 (2), 141-158).
- García Hoz, V. (1953). *Vocabulario Usual, Común y Fundamental*. Madrid: CSIC. Instituto de Pedagogía.
- García Hoz, V. (1976a). Test de vocabulario. En V. García Hoz *Manual de test para la EGB* (pp. 27-40). Madrid: Escuela Española.
- García Hoz, V. (1976b). *Vocabulario General de Orientación Científica y sus estratos*. Madrid: Instituto de Pedagogía, CSIC.
- García Hoz, V. (1977). *Estudios experimentales sobre vocabulario*. Madrid: CSIC. Instituto de Pedagogía S. José de Calasanz.
- García Hoz, V. (1999). El vocabulario común y su utilización didáctica. *Bordón*, 51, 2, 141-158.
- García M., J. A., Martín, J. I., Luque, J. L. e Santamaría, C. (1995) *Comprensión y adquisición de conocimientos a partir de textos*. Madrid: Siglo XXI.
- García Vidal, J. (1999). *La evaluación predictiva de la iniciación lectora*. Tese de doutoramento. Universidad de Sevilla. Facultad de CC. de la Educación.
- García, C. e González, M. (Dir.) (1997). *Diccionario da Real Academia Galega*. Vigo: Edicións Xerais. A Coruña: Galaxia.
- Genesee, F. (1989). Early bilingual development: one language or two?. *Journal of Child Language*, 16, 161-179.
- Genouvrier, E. e Peytard, J. (1970). *Linguistique et enseignement du français*. Paris: Larousse.
- Gentner, D. (1975). Evidence for the psychological reality of semantic components. The verbs of possession. En D. A. Norman e D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 211-246). San Francisco: W.H. Freeman and Co.
- Giammatteo, M. e Basualdo, M. (2003). Desempeño léxico y producción escrita. Análisis y comparación de resultados de una experiencia aplicada en la escuela secundaria. *Cultura y Educación*, 15 (1), 29-46.
-

- Gilabert, R. (1995). Enseñanza de estrategias para la inferencia del significado de las palabras. *Infancia y Aprendizaje*, 72, 139-152.
- Gilabert, R. e Vidal-Abarca, E. (1993). ¿Podemos favorecer el aprendizaje autónomo del significado de palabras? *Comunicación, Lenguaje y Educación*, 18, 81-90.
- Gómez F., D. E. (Dir.) (1988). *Implantación de programas de lenguaje compensatorio en el contexto bilingüe de Galicia*. Madrid: CIDE.
- Gómez G., X. e Torres, A. (2005). *Vocabulario xurídico administrativo galego castelán*. Vigo: Universidade de Vigo (Servizo de publicacións).
- González Lorenzo, M. (1981). *La práctica bilingüe gallego-castellano en el desarrollo cognitivo y verbal*. Santiago de Compostela: Universidade de Santiago de Compostela. Tese de doutoramento.
- González Lorenzo, M. (1985). *Bilingüismo en Galicia. Problemas y alternativas*. Santiago de Compostela: Servicio de Publicacións da Universidade.
- Gopnik, A. e Meltzoff, A. N. (1992). Categorization and naming: basic-level sorting in eighteen-month-olds and its relation to language. *Child Development*, 63, 1091-1103.
- Gough, Ph. B., Hoover, W. A. e Peterson, C. L. (1996). Some Observations on a Simple View of Reading. En C. Cornoldi e J. Oakhill, *Reading Comprehension Difficulties. Processes and Intervention (pp. 1-13)*. Mahwah, NJ : Erlbaum.
- Goulden, R., Nation, P. e Read, J. (1990). How Large Can a Receptive Vocabulary Be?. *Applied Linguistics*, 11 (4), 341-363.
- Graves, M. F. (1986). Vocabulary learning and instruction. En E. Rothkopf (Ed.), *Review of research in education (pp. 49-89)*. Washington, DC: American Educational Research Association.
- Graves, M. F. (1987). The roles of instruction in fostering vocabulary development. En M. G. McKeown e M. E. Curtis (Eds.), *The Nature of Vocabulary Acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Green, D. M. e Swets, J. A. (1966). Signal detection theory and psychophysics. *Psychological Bulletin*, 75, 424-429.
- Greenwald, E. A., Persky, H. R., Campbell, J. R. e Mazzeo, J. (1999). *The NAEP 1998 writing report card for the nation and the states*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Gutiérrez-Calvo, M. (1999). Inferencias en la comprensión del lenguaje. En M. de Vega e F. Cuetos (Coords.), *Psicolingüística del español (pp. 231-270)*. Madrid: Trotta.
- Gutiérrez-Calvo, M. e Castillo, M. D. (1997). Mood Congruent Bias in Interpretation of Ambiguity: Strategic Processes and Temporary Activation. *Quarterly Journal of Experimental Psychology*, 50A, 163-182.

-
- Gutiérrez-Calvo, M., Eysenck, M. W. e Estévez, A. (1994). Ego-Threat Interpretive Bias in Test Anxiety: On-Line Inferences. *Cognition and Emotion*, 8, 127-146.
- Gutiérrez-Calvo, M., Eysenck, M. W. e Castillo, M. D. (1997). Interpretative Bias in Test Anxiety: The Time Course of Predictive Inferences. *Cognition and Emotion*, 11, 43-63.
- Haber, A. e Runyon, R. P. (1973). *Estadística general*. México DF: Fondo Educativo Interamericano.
- Hall, W. S. e Tirre, W. C. (1979). *The communicative environment of young children: social class, ethic and situational differences*. Urbana, IL: University of Illinois, Center for the Study of Reading (Inf. Tec. Nº 125), (ERIC ED 170.788).
- Halpin, G., Simpson, R. e Martin, S. L. (1990). An investigation of racial bias in the Peabody Picture Vocabulary Test-Revised. *Educational and psychological Measurement*, 50 (1), 183-189.
- Harris, R. (2000). *Rethinking writing*. London: Continuum.
- Harris, T. L. e Hodges, R. E. (Eds.) (1995). *The Literacy Dictionary: The Vocabulary of Reading and Writing*. Newark, DE: International Reading Association.
- Harris, J. e Wilkinson, J. (1986). Reading Children's Writing. *A Linguistic View*. London: Allen e Unwin Publishers Ltd.
- Harrison, C., Bayley, M. e Deward, A. (1998). Responsive reading assessment: is postmodern assessment possible?. En C. Harrison e T. Salinger (Eds.), *Assessing reading: theory and practice (1-20)*. London: Routledge.
- Hasan, R. (1985). The texture of a text. En M. A. K. Halliday e R. Hasan, *Language, context and text: aspects of language in a social semiotic perspective (pp. 70-96)*. Oxford: Oxford University Press.
- Hayes, F. B. e Martin, R. P. (1986). Effectiveness of the PPVTR in the screening of young gifted children. *Journal of Psychoeducational Assessment*, 4 (1), 27-33.
- Hockett, C. F. (1958). *A course in Modern Linguistics*. New York: The MacMillan Co.
- Jiménez, B. (1986). *Didáctica del vocabulario*. Barcelona: Humanitas.
- Jiménez, J. e Artiles, C. (1989). *Cómo prevenir y corregir las dificultades en el aprendizaje de la lecto-escritura*. Madrid: Síntesis.
- Jiménez, J. e Artiles, C. (1990). Factores predictivos del éxito en el aprendizaje de la lecto-escritura. *Infancia y Aprendizaje*, 49, 21-37.
- Johnson, D. D. (2000). Just the Right Word: Vocabulary and Writing. En R. Indrisano e J. R. Squire (Eds.), *Perspectives on Writing. Research, Theory and Practice (pp. 162-186)*. Newark, Delaware: IRA.
-

- Johnson, D. D.; Moe, A. J. e Baumann, J. F. (1983). *The Ginn word book for teachers: A basic lexicon*. Lexington, MA: Ginn and Company.
- Johnston, P.H. (1989). *La evaluación de la comprensión lectora*. Madrid: Aprendizaje-Visor.
- Jones, S. S., Smith, L. B. e Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, 62, 499-516.
- Juilland, A. e Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words*. London: Mouton & Co.
- Justicia, F. (1983). *Determinación y análisis del vocabulario usual del niño de 6 a 10 años en Andalucía Oriental*. Granada: Tese Doutoramento. ICE de la Universidad.
- Justicia, F. (1985a). El vocabulario del niño en los primeros cursos de EGB. *Enseñanza*, 3, 29-39.
- Justicia, F. (1985b). *Vocabulario usual del niño en el Ciclo Inicial y el Ciclo Medio de la EGB*. Granada: ICE de la Universidad.
- Justicia, F. (1993). Metodología en la estimación del vocabulario infantil. *Revista Española de Lingüística Aplicada*, Año IX, 125-136.
- Justicia, F. (1995). *El desarrollo del vocabulario: diccionario de frecuencias*. Granada: Universidad de Granada.
- Katz, J. J. e Fodor, J. A. (1976). *Estructura de una teoría semántica*. Madrid: Siglo XXI.
- Kintsch, W. (1998). *Comprensión. A Paradigm for Cognition*. Cambridge: Cambridge University Press.
- Kirk, S. A., McCarthy, J. J. e Kirk, W. D. (1989). *Test ITPA*. Madrid: TEA.
- Kline, P. (1986): *A Handbook of Test Construction. Introduction to Psychometric Design*. London: Methuen & Co. Ltd.
- Kress, G. (2005). Gain and losses: New forms of texts, knowledge and learning. *Computers and Composition*, 22, 5-22.
- Laufer, B. e Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16 (3), 307-322.
- Laufer, B. e Paribakht, T. S. (1998). The relationship Between Passive and Active Vocabularies: Effects of Language Learning Context. *Language Learning*, 48, 3, 365-391.
- Lázaro Carreter, F. (1981). *Diccionario de términos filológicos*. Madrid: Gredos (3ª).
- León, O. e Montero, I. (1997). *Diseño de investigaciones. Introducción a la lógica de la investigación en Psicología y Educación*. Madrid: McGraw-Hill.

-
- Lecuona, M. P. (1988). *El vocabulario usual del preescolar salmantino*. Salamanca: Universidad de Salamanca. Tese de Doutoramento.
- Lewis, D. G. (1971). *The Anaysis of Variance*. Manchester: Manchester University Press.
- Lieven, E. V., Pine, J. M. e Dresner, H. (1992). Individual differences in early vocabulary development: redefining the referencial-expressive distinction. *Journal of Child Language*, 19, 287-310.
- Livingstone, S. A. (1972). A criterion-referenced applications of classical test theory. *Journal of Educational Measurement*, 9, 13-26.
- López F., R. (1990). Sobre os dicionarios galegos. *A Trabe de Ouro*, 3, 395-409.
- López Morales, H. (1984). *Enseñanza de la lengua materna*. Madrid: Playor.
- Lorge, I. e Chall, J. (1963). Estimating the Size of Vocabulary of Children and Adults: An Analysis of Methodological Issues. *Journal of Experimental Education*, 32, 147-157.
- Luria, A. R. (1978). *Lenguaje y desarrollo intelectual en el niño*. Madrid: Pablo del Río.
- Luria, A. R. (1980). *Conciencia y lenguaje*. Madrid: Pablo del Río.
- Lyons, J. (1980). *Semántica*. Barcelona:Teide.
- Malkiel, Y. (1970). Genetic Analysis of Word. *Current Trends in Linguistics*, 3, 305-364.
- Manning, M. (1999). Helping Word Grow. *Teaching Prek-8*, 29 (4), 103-105.
- Maratsos, M. (1990). Are actions to verbs as objects are to nouns? On the differential semantic bases of form, class, category. *Linguistics*, 28, 1351-1379.
- Markman, E. M. (1989). *Categorization and naming in children. Problems in induction*. Cambridge, MA: MIT Press.
- Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, 92, 199-227. (Reeditado en L. Gleitman e B. Landau (Eds.) (1996), *The Acquisition of the Lexicon*. Cambridge, MA: MIT Press).
- Márquez Villegas, L. (1975). *Vocabulario del español hablado*. Madrid: SGEL.
- Martínez, A. (2004). O mundo rural hoxe. *Que quede quen o conte (suplemento)*, Marzo.
- Martinet, J. (1966) Le mot . En *Problèmes du language* (pp.39-53). París: Gallimard.
- Martinet, J. (1972). *La lingüística. Guía alfabética*. Madrid: Anagrama.
- McCarthy, D. (1996). *Test MSCA*. Madrid:TEA.
- McShane, J. (1980). *Learning to talk*. Cambridge: Cambridge University Press.

- Meara, P. M. (1978). Learner's word association in French. *The Interlanguage Studies Bulletin*, 3, 2, 192-211.
- Meara, P. M. (1987). *Vocabulary in a Second Language Vol. 2*. London: CILT.
- Meara, P. M. (1990). Some notes on the Eurocentre's Vocabulary Test. En J. Tommola (ed.) *Foreign Language Comprehension and Production* (pp. 103-113). Turku: AFinLA.
- Meara, P. M. (1991). *Scoring a Yes/No Vocabulary Test*. Draft for discussion. Swansea: Centre for Applied Language Studies. University College.
- Meara, P. M. (1993a). *EFL Vocabulary Test*. Swansea: Centre for Applied Language Studies. University College.
- Meara, P. M. (1993b). The Bilingual Lexicon and the Teaching of Vocabulary. In R. Schruder e B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 279-297). Amsterdam: John Benjamins.
- Meara, P. M. (1994). The complexities of simple vocabulary test. En F. G. Brinkman, J. A. von der Schee e M. C. Schouten-van Parreren (Eds), *Curriculum research: different disciplines and common goals* (pp. 15-28). Amsterdam: Instituut voor Didactiek eu Onderwijspraktijk, Vrije Universiteit.
- Meara, P. M. (1996). *The Vocabulary Knowledge Framework*. Swansea: Centre for Applied Language Studies. University College (On-line <http://www.swansea.ac.uk/cals/vlib.html>).
- Meara, P.M. (2005). *X_LEX: the Swansea Vocabulary Levels Test. V2.05*. Swansea: Lognostics.
- Meara, P. M. e Buxton, B. (1987). An alternative to multiple choice vocabulary test. *Language Testing*, 4 (2), 142-154.
- Meara, P. M., Huibregtse, I. e Admiraal, W. (2002). Scores on a yes-no vocabulary test: correction for guessing and response style. *Language Testing*, 19 (3), 227-245.
- Meara, P. M. e Jones, G. (1988). Vocabulary size as a placement indicator. En P. Grunwell (Ed.), *Applied Linguistics in Society* (pp. 80-87). Papers from Annual Meeting of the British Association for Applied Linguistics, Nottingham, UK: CILT.
- Meara, P. M. e Jones, G. (1990). *The Eurocentres Vocabulary Size Test. 10 KA*. Zurich: Eurocentres.
- Meara. P. M., Ligthbown, P. M. e Halter, R. (1994). The effect of cognates on the applicability of YES/NO vocabulary test. *Canadian Modern Language Review*, 50, (2), 296-311.
- Meara. P. M. e Rodríguez Sánchez, I. (1993). *Matrix models of Vocabulary Acquisition: an empirical assessment*. Ottawa: CREAL, Symposium on Vocabulary Research (On-line <http://www.swansea.ac.uk/cals/vlib.html>).

-
- Medo, M. A. e Ryder, R. J. (1993). The Effects of Vocabulary Instruction on Readers' Ability to Make Causal Connections. *Reading research and Instruction*, 33 (2), 119-134.
- Meisel, J. M. e Mahlau, A. (1988). La adquisición simultánea de dos primeras lenguas. Discusión general e implicaciones para el estudio del bilingüismo en Euzkadi. *Iberoamericana*, 2/3, 5-34.
- Michel, J. (1972). *The Problem of Time: Some Techniques for Teaching Vocabulary*. ERIC Reports, 27, New York. Citado en A. Suárez (1987).
- Miller, G. e Gildea, P. (1987). How Children Learn Words. *Scientific American*, 257 (3), 94-99.
- Milton, J. e Meara, P. M. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics*, 107-108, 17-34.
- Molina, S. e García, E. (1984). *El éxito y el fracaso escolar en la EGB*. Barcelona: Laia.
- Muñiz, J. (2000): *Teoría clásica de los tests*. Madrid: Pirámide.
- Nagy, W. E. (1988). *Teaching Vocabulary to Improve Reading Comprensión*. Newark, Delaware: IRA.
- Nagy, W. E. (1997). On the role of context in first and second-language vocabulary learning. En N. Schmitt e M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 64-83). Cambridge, U K: Cambridge University Press.
- Nagy, W. E. e Anderson, R. C. (1984). How many words are there in printed school English?. *Reading Research Quarterly*, 19, 304-330.
- Nagy, W. E. e Herman, P. A. (1987). Breadth and Depth of Vocabulary Knowledge: Implications for Acquisition and Instruction. En M. G. McKeown e M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (17-35). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Nagy, W. E., Herman, P. A. e Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233-253.
- Nagy, W. E. e Scott, J.A. (2000). Vocabulary processes. En M. I. Kamil, P.B. Mosenthal; P.D. Pearson e R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 269-284). Mahwah, NJ: Erlbaum.
- Nation, P. (1990). *Teaching and Learning Vocabulary*. Boston: Heinle & Heinle.
- Nation, P. (1993a). Using dictionaries to estimate vocabulary size: essential but rarely followed procedures. *Language Testing*, 10 (1), 27-40.
- Nation, P. (1993b). Measuring Readiness for Simplified Material: A Test of the First 1.000 Words of English. En M.L. Tickoo (Ed.), *Simplification: Theory and Application. Antology Series 31* (pp. 193-203). (ERIC ED 371.585).
-

- Nation, P. e Ellis, B. (1993). *Vocabulary Growth*. Citado en P. Nation (1993a).
- National Institute of Child Health and Human Development (2000). *The report of the National Panel. Teaching children to read: An evidence-based assessment of the scientific literature on reading and its implications for reading instruction*. Washington, DC: U.S. Government Printing Office.
- Nelson, K. (1988a). *El descubrimiento del sentido*. Madrid: Alianza.
- Nelson, K. (1988b). Constraints on word learning. *Cognitive Development*, 3, 221-246.
- New Standards (1995). *Performance Standards English Language Arts, Mathematics, Science and Applied Learning, Vols. 1, 2 and 3. Consultation Drafts*. Washington, DC: National Center for Education and the Economy.
- O'Brien, D. G. (1986). A test of three positions posed to explain the relation between word knowledge and comprehension. *National Reading Conference Yearbook*, 35, 81-86.
- Oleron, P. (1981). *El niño y la adquisición del lenguaje*. Madrid: Morata.
- Palmer, F. R. (1983). *Semantics*. Cambridge: Cambridge University Press.
- Pandolfi, A. M. e Herrera, M. O. (1985). Análisis léxico-semántico en el niño menor de tres años. *Revista de Lingüística Teórica y Aplicada*, 23, 121-131.
- Pastora, J. F. (1990). *El vocabulario como agente de aprendizaje*. Madrid: La Muralla.
- Pastorre, R. E. e Scheirer, C. J. (1974). Signal Detection Theory: considerations for general applications. *Psychological Bulletin*, 81, 945-958.
- Pease, D. e Berko, J. (1985). Gaining Meaning: Semantic Development. En J. Berko (Ed.), *The development of Language (pp. 103-138)*. Columbus: Charles E. Merrill.
- Peraíta, H. (1986). Desfase entre la producción y la comprensión del léxico de niños de 24 a 36 meses de edad. *Infancia y Aprendizaje*, 35/36, 11-24.
- Pérez González, J. (1981). Vocabulario. En J. Luján (Dir.), *Elaboración de instrumentos para la evaluación de aspectos básicos del rendimiento escolar en 8º curso de EGB (pp. 30-42)*. Madrid: MEC. Estudios y Experiencias Educativas. Serie Evaluación.
- Pérez P., M. e Castro, J. (1990). Adquisición da linguaxe en preescolar e no ciclo inicial. *Cadernos de Psicoloxía*, 8, 34-42.
- Pérez Vilariño, J. (1979). Lengua y discriminación escolar en Galicia. *Jornadas de bilingüismo (pp. 213-229)*. Valencia: Universidad de Valencia.
- Pérez Vilariño, J. (1981). Ámbitos sociales y usos lingüísticos entre los escolares gallegos. *Revista de Educación*, 268, 111-130.
- Pergnier, M. (1986). *Le mot*. París: Presses Universitaires de France.

-
- Peterson, C. L. (1993). *Background knowledge and the decomposition of literacy in skilled adult readers*. Austin: University of Texas. Unpublished doctoral dissertation. Citado en Ph. B. Gough, W.A. Hoover e C.L. Peterson (1996).
- Pikulski, J. e Tobin, W. (1989). Factors associated with longterm reading achievement of early readers. *National Reading Conference Yearbook*, 38, 123-133.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: Harper Perennial.
- Qian, D. D. (1999). Assessing the Roles of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension. *Canadian Modern Language Review*, 56 (2), 282-307.
- Quilis, A. e Fernández, J. A. (1979). *Curso de fonética y fonología españolas para estudiantes angloamericanos*. Madrid: CSIC.
- Ramírez, A. G. (1973). Bilingüismo y centros piloto de EGB en Galicia. *Iª Jornadas de Trabajo sobre Bilingüismo*. Santiago de Compostela: ICE da Universidade.
- Raphael, T. E., Freebody, P., Fritz, M., Mayers, A. C. e Tirre, W. C. (1980). *Contrasting the effects of some text variables on comprehension and ratings of comprehensibility*. Ponencia presentada no Congreso de American Educational Research Association, Boston. Citado en P. H. Johnston (1989).
- Read, J. (1987). *Towards a deeper assessment of vocabulary knowledge*. Sidney: Paper presented at the 8th World Congress of Applied Linguistics. ERIC Clearinghouse on Languages and Linguistics (ED 301.048).
- Read, J. (1990). *The Word Associates Test: A measure of quality of vocabulary knowledge*. Documento de traballo non publicado. Citado en Vives Boix (1995).
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10 (3), 355-371.
- Read, J. (1995). Validating the Word Associates Format as a measure of depth of vocabulary knowledge. *Traballo presentado no 17º Language Testing Research Colloquium*, Long Beach, California. Citado en J. Cummins (2002).
- Read, J. (1997). Assessing vocabulary in a second language. En C. Clapham e D. Corson (Eds.), *Language Testing and Assessment. Vol 7. Encyclopedia of Language and Education* (pp. 99-107). Dordrech : Kluwer Academic Publishers.
- Real Academia Española (1987). *Diccionario de la lengua española*. Madrid. (20.ª edic.).
- Real Academia Española (2001). *Diccionario de la lengua española*. Madrid. (22.ª edic.).
- Reichler-Béguelin, M. J. (1992). Perception du mot graphique dans quelques systèmes syllabiques et alphabétiques. *Lalies*, 10. Publications de la Sorbonne Nouvelle.
- Rey-Devobe, J. (1976). Problèmes de sémantique lexicale. En B. Pottier (Ed.), *Sémantique et logique* (pp. 167-182). Paris: Jean Pierre Delarge.

- Reznick, J.S. e Goldsmith, L. (1989). A multiple form word production checklist for assessing early language. *Journal of Child Language*, 16, 91-101.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quaterly*, 10, 77-89.
- Robbins, C. e Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, 86 (1), 54-64.
- Rodríguez Bou, I. (1952). *Recuento del vocabulario español*. San Juan de Puerto Rico: OEA.
- Rojo, G. (1981). En torno a las actitudes lingüísticas de los profesores de EGB en Galicia. *Revista de Educación*, 268, 131-155.
- Rondal, J. A. (1982). *El desarrollo del lenguaje*. Barcelona: Ed. Médica y Técnica.
- Rosch, E. e Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Rubal, X. (Coord.) (2002). *Léxico de métodos de investigación e diagnóstico en educación*. Santiago de Compostela: Universidade de Santiago de Compostela (Servizo de Publicacións e Intercambio Científico).
- Rubal, X. e Rodríguez, M. A. (1987). *O galego no ensino público non universitario*. Santiago de Compostela: Consello da Cultura Galega.
- Ruddell, M. R. (1994). Vocabulary knowledge and comprehension: A comprehension-process view of complex literacy relationships. En R. B. Ruddell, M. R. Ruddell e H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 414-447). Newark, DE: International Reading Association.
- Rust, J. e Golombok, S. (1989). *Modern Psychometrics. The science of Psychological Assessment*. New York: Routledge.
- Sánchez, E. (1993). *Los textos expositivos. Estrategias para mejorar su comprensión*. Madrid: Aula XXI-Santillana.
- Sánchez, E. (1996). El todo y las partes: una crítica a las propuestas del lenguaje integrado. *Cultura y Educación*, 1, 39-54.
- Santos, M. A., Cajide, J., Lorenzo, M^a. M. e Porto, A. M^a. (2002). A lingua galega no ensino primario: datos de hoxe para unha análise de alcance. En M^a. X. Bugarín, J. Cajide, A. Dosil, X. G. Ferreiro, M. González e M. A. Santos (Eds.), *Actas da VIII Conferencia Internacional de Linguas Minoritarias* (pp. 609-613). 22-24 novembro de 2001. Santiago de Compostela: Xunta de Galicia.
- Saramago, J. (2003). *El hombre duplicado*. Madrid: Santillana Ediciones Generales.
- Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology*, 81 (1), 101-108.

-
- Scarborough, H. S. (1991). Early syntactic development of dyslexic children. *Annals of Dyslexia*, 41, 207-220.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11 (2), 129-158.
- Schmitt, N. (1998). Tracking the Incremental Acquisition of Second Language Vocabulary: A Longitudinal Study. *Language Learning*, 48, 2, 281-313.
- Schonell, F. J., Meddleton, I. G. e Shaw, B. A. (1956). *A Study of Oral Vocabulary of Adults*. Brisbane: University of Queensland Press. Citado en P. Nation (1993b).
- Schoonen, R., Hulstijn, J. e Bossers, B. (1998). Metacognitive and Language-Specific Knowledge in Native and Foreign Language Reading Comprehension: An Empirical Study Among Dutch Students in Grades 6, 8 and 10. *Language Learning*, 48 (1), 71-106.
- Sebastián, N., Martí, M. A., Cuetos, F. e Carreiras, M. (2000). *LEXEP. Léxico informatizado del español*. Barcelona: Ediciones de la Universitat de Barcelona.
- Seisdedos, N., Cruz, M^a. V. de la, Cordero, A. e González, M. (1990). *Test TEA (6.ª edic.)*. Madrid: TEA.
- Seoane, J. e Rechea, C. (1976). Unidad didáctica 5. En J. Seoane e C. Rechea, *Psicometría (pp. 55-64)*. Madrid: UNED.
- Serra, M. e Serrat, E. (1996). Diferencias individuales en la adquisición del léxico: revisión de la producción y organización del léxico y de su procesamiento inicial. En M. Pérez (Ed.), *Estudios sobre la adquisición del castellano, catalán, eusquera y gallego (pp. 301-315)*. Santiago de Compostela: Universidad de Santiago de Compostela.
- Serra, M., Serrat, E., Solé, R., Bel, A. e Aparici, M. (2000). *La adquisición del lenguaje*. Barcelona: Ariel Psicología.
- Shaughnessy, M. F. e Evans, R. (1986). Word/world knowledge: Prediction of college GPA. *Psychological Reports*, 59 (3), 1147-1150.
- Sierra Bravo, R. (1991). *Técnicas de investigación social: teoría y ejercicios*. Madrid: Paraninfo.
- Smith, C. B. (1997a). *Word List for a Spelling Program*. Paper presented at the Annual International Dyslexic Conference. Minneapolis, Nov. 14-15. (ERIC ED 412.507).
- Smith, C. B. (1997b). *Vocabulary Instruction and Reading Comprehension*. ERIC Clearinghouse on Reading English and Communication. Bloomington, IN. (ED 412.506).
- Smith, E. E., Shoben, E. J. e Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214-241.
-

- Smith, M. E. (1926). An investigation of the development of the sentence and the extent of vocabulary in young children. *Studies in Child Welfare*, Vol. 3, nº 5. Iowa City: University of Iowa. Citado en Ph. S. Dale (1976).
- Smith, M.K. (1941). Measurement of the size of general English vocabulary through the elementary grades and high school. *Genetic Psychological Monographs*, 24, 311-345. Citado en D. D. Johnson (2000).
- Snow, C. E., Burns, M. S. e Griffin, P. (Eds.) (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Sobrado, L. e Porto, A. (Corrds.) (2000). *Glosario de Orientación Escolar e Profesional*. Santiago de Compostela: Universidade de Santiago de Compostela (Servizo de Publicacións e Intercambio Científico).
- Solomon, J. e Rhodes, N. (1996). Assessing academic language: Results of a survey. *TESOL Journal*, 5 (4), 5-8.
- Spillich, G. J., Vesonder, G. T., Chiesi, H. L. e Voss, J. F. (1979). Texts processing of domain related information for individuals with high and low domain knowledge. *JVLVB*, 18, 275-290.
- Spring, C. e Prager, J. (1992). Teaching community-college students to follow the train of thought in expository text. *Reading and Writing: An Interdisciplinary Journal*, 4, 33-54.
- Stahl, S. (1991). Defining the Role of prior Knowledge and vocabulary in Reading Comprehension: The retiring of Number 41. *Journal of Reading Behavior*, Vol. 23 (4): 487-508.
- Stahl, S. A. e Fairbanks, M. (1986). The effects of vocabulary instruction: A model based on meta-analysis. *Review of Educational Research*, 56, 72-110.
- Stallman, A. C. (1991). *Learning vocabularies from context: Effects of focusing attention on individual words during reading*. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign. Citado en W.E. Nagy (1997).
- Stanovich, K. E. (1986). Mathew's effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21 (3), 360-407.
- Stanovich, K. E. e Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory and Cognition*, 20, 51-60.
- Stanovich, K. E., Nathan, R. G. e Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading Research Quarterly*, 21 (3), 267-283.

-
- Sternberg, R. (1987). Most vocabulary is learned from context. En M.G. McKeown e M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Erlbaum.
- Stoel-Gammon, C. e Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, 11, 247-271.
- Stuart, M., Masterton, J. e Dixon, M. (2000). Spongelike acquisition of sight vocabulary in beginning readers? *Journal of Research in Reading*, 23 (1), 12-27.
- Suárez, A. (1987). *Papel desempeñado por factores ortográficos en la "codificación" fonológica de palabras inglesas por parte de hablantes de español nativos*. Tese de doutoramento. Madrid: Universidad Complutense.
- Suárez, A. (2000). *Iniciación escolar a la escritura y la lectura. Diseño de programas adaptados a la diversidad*. Madrid: Pirámide.
- Suárez, A., Seisdedos, N. e Meara, P. (1998). *Test EVOCA*. Madrid: TEA.
- Suárez García, J. (1999). Review: Diccionario de frecuencias de las unidades lingüísticas del castellano de J. R. Alameda e F. Cuetos (1995) (On-line <http://www.swansea.ac.uk/cals/vlib.html>).
- Swaborn, M.S.L. e de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69, 261-286.
- Tan, A. e Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology*, 89, 276-288.
- Tarnowski, K.J. e Kelly, P.A. (1987). Utility of the PPVT-R for pediatric intellectual screening. *Journal of Pediatric Psychology*, 12 (4), 611-614.
- Tejedor, F. J. (1984). *Análisis de varianza aplicado a la investigación en Pedagogía y Psicología*. Madrid: Anaya.
- Teijido, M^a. C., Pérez, M^a. L. e Fraga, I. (2001). *Diccionario de frecuencias léxicas do galego escrito para nenos (LEXIGAL Primaria)*. Santiago de Compostela: Servizo de Publicacións da USC. Editado en CD-Rom.
- Thorndike, E. L. (1924). The vocabularies of school pupils. En J. Carelton Bell (Ed.). *Contribution to education* (pp. 69-76). New York: World Book Co. Citado en P. Nation (1993a).
- Thorndike, E. L. e Lorge, I. (1944/1963). *The teacher's word book of 30000 words*. New York: Teacher's College, Columbia University.
- Tierney, R. J. e Cunningham, J. W. (1984). Research on teaching reading comprehension. En P.D. Pearson (Ed.), *Handbook of reading research* (pp. 609-655). New York: Longman.

- Tierney, R. J. e Shanahan, T. (1991). Research on the reading-writing relationship: Interaction, transactions and outcomes. En R. Barr, M. L. Kamil, P. Moshental e P. D. Pearson (Ed.), *Handbook of reading research. Vol 2* (pp. 246-280). New York: Longman.
- Tillinghast, B. S., Morrow, J. E. e Uhlig, G. E. (1984). Validity of the PPVT-R using elementary school academic rades as criteria. *Psychology a Quaterly Journal of human Behavior*, 21 (2), 43-47.
- Ullmann, S. (1972). *Semántica: introducción a la semántica del significado*. Madrid: Aguilar.
- Valle A., F. (1991). *Psicolingüística*. Madrid: Morata.
- Vance, B., West, R. e Kutsick, K. (1989). Prediction of Wechsler Preschool and primary scale of Intelligence IQ scores for preschool children using Peabody Picture Vocablary Test-Revised and the Expressive One Word Picture Vocabulary Test. *Journal of Clinical Psychology*, 45 (4), 642-644.
- Vega, M. de e Cuetos, F. (Coords.) (1999). *Psicolingüística del español*. Madrid: Trotta.
- Verhoeven, L.T. (1990). Acquisition of reading in a second language. *Reading Research Quarterly*, 25, 90-114.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press. (Tradución: *Pensamiento y lenguaje*, 1992. Buenos Aires: Pléyade).
- Vives Boix, G. (1995). *The development of a measure of lexical organitation: the association vocabulary test*. Swansea: Tese Doutoramento. University College of Swansea. University of Wales.
- Watts, A. F. (1946). *The Language and Mental Development of Children*. London: D.C. Heath. Citado en García Hoz (1953).
- Wechsler, D. (1989, 1993). *Test WISC*. Madrid:TEA.
- Wesche, M. e Paribakht, T. S. (1993). Assessing Vocabulary Knowledge: Depth versus Breadth. *Paper presented to the 15th Language Testing Colloquium*. Arnhem.
- Wesche, M. e Paribakht, T. S. (1994). Enhancing Vocabulary Acquisition trough Reading: A Hierarchy of Text-Related Exercise Types. *Paper presented at the Annual Meeting of the American Association for Applied Linguistics*. Baltimore, MD. (ERIC ED 369.291).
- Wesche, M. e Paribakht, T. S. (1996). Assessing Second Language Vocabulary Knowledge: Depth versus Breadth. *The Canadian Modern Language Review*, 53, 1, 13-40.
- West, M. (1953). *A General Sevice List of English Words*. London: Longman, Green an Co. Citado en L. Bauer e P. Nation (1993).

-
- White, T. G., Wayne, H. S. e Graves, M. F. (1989). YES/NO Method of Vocabulary Assesment: Valid for Whom and Useful for What?. In S. McCormick e J. Zutell (Eds.), *Cognitive and Social Perspectives for Literacy Research and Instruction*. Chicago: Thirty-eight Yearbook. National Reading Conference Tests.
- Williams, H. M. (1932). Some problems of sampling in vocabulary test. *Journal of Experimental Education*, 1, 131-133. Citado en Nation (1993a).
- Wray, D. e Lewis, M. (2000). *Aprender a leer y escribir textos de información*. Madrid: Morata.
- Xue, G. e Nation, P. (1984). A university word list. *Language Learning and Communication*, 3, 2, 215-229.
- Xunta de Galicia (1999). *Poboación escolar de Galicia (Curso académico 1998/1999)*. Información proporcionada polo Servicio de Planificación e Sistemas de Información da Consellería de Educación e Ordenación Unversitaria.
- Zimmerman, J., Broder P., Shaughnessy, J. e Underwood, B. (1977). A recognition test vocabulary using signal-detection measures, and some correlates of word and non-word recognition. *Intelligence*, 1, 5-31.
- Zucker, S. e Riordan, J. (1988). Concurrent validity of new and revised conceptual language measures. *Psychology in the Schools*, 25 (3), 252-256.

Anexos

Anexo I

Instrucións ofrecidas ós xuíces para valorar a dificultade de aprendizaxe das palabras

A tarefa que che solicito consiste en clasifica-las palabras das listas de acordo co GRAO DE DIFICULTADE DE APRENDIZAXE que presenta cada unha das palabras segundo o teu xuízo.

Para realizala, podes proceder do seguinte xeito:

Primeiro, dalle un vistazo a tódalas palabras das listas co fin de obter unha impresión xeral da dificultade das mesmas.

A continuación, lendo xa as palabras máis pausadamente, detrás de cada unha delas escribe un 1, un 2, un 3 ou un 4, conforme ó código que che indico seguidamente:

- 1 = unha palabra das máis fáciles de aprender.
- 2 = unha palabra un pouco máis difícil de aprender que a marcada cun 1.
- 3 = unha palabra máis difícil de aprender que a marcada cun 2, pero menos difícil que a marcada cun 4.
- 4 = unha palabra das máis difíciles de aprender.

Emite un xuízo rápido, non se trata de deterse demasiado pensándoo.

Ó final, procura que haxa aproximadamente o mesmo número de palabras en cada un dos catro grupos (unhas 500 en cada un).

Se modifica-lo número asignado a unha palabra, trata de que a resposta definitiva se lea con claridade.

Se che resulta excesivamente complicado, podes reparti-la lista con outra persoa das mesmas características profesionais ca ti.

Anexo II

Táboas coas pseudopalabras incluídas nos 10 subtets e o número de orde que ocupan en cada un

TEST nº 1	TEST nº 2	TEST nº 3	TEST nº 4	TEST nº 5
presuntar	arrancadoiro	rabaliza	labiza	soreiro
rendibilizar	cebrería	rufalco	fanego	edulceración
radiosónico	aramalla	astroide	zaparallada	ledicioso
moarí	cánodo	desimpactante	ornicultura	canceiro
mallarada	desfanar	funxido	felinio	sárrico
nárbaro	carcidume	áuleo	bocellar	merica
bracicar	trolo	aleísmo	quecemanchas	cinleiro
croquela	esloca	postada	recolmar	alcén
grechar	alfaiote	baldo	guilleiro	incepente
solegar	deslistar	carraxeiro	chalano	consurcar
rastrelar	serativo	maricidio	róbado	moileiro
cediña	impermanente	anollar	desbanar	anerxia
eslomecer	zaspán	tragalinguas	desdexar	investidiario
larpento	teigume	embanelar	tubullo	fusto
esbarecer	pasabrisas	cónchego	aromalía	seneridade
colomego	trepento	xordinte	calenán	caletrán
embornar	involveinte	desbartar	ouleiro	liañar
colzar	refulgar	esbullar	atego	helanismo
esmoular	acelgar	desbrazamento	avanceiro	indirixible
raciño	receitamento	pingoreteiro	samergo	labreiro

TEST nº 6	TEST nº 7	TEST nº 8	TEST nº 9	TEST nº 10
papadeiro	intervidura	fanatidade	raspadela	fragulla
maristada	reluscencia	chanella	apestadeira	tullereiro
bretoniano	ivar	naturisto	tremulecente	artén
rexarcir	imbanación	cernidor	follieiro	prosbentos
fervencia	zouneiro	xermadólogo	frisada	tendinismo
esfingulación	piláculo	esnourar	quereza	teimura
abieiro	despicaz	pardizo	larugar	devanzar
alivar	repositar	carrucén	tabanismo	molella
amioeiro	nuril	hílico	aranceira	caledo
franel	engolar	perliscente	lerdeiro	súpula
impresentar	fiteira	proxulgar	loumado	desnilar
entear	alcasán	laniño	barrentar	demalar
lunoso	lagoar	auniense	publizar	descéntrico
presuntivo	asebullón	hemógrafo	mounelo	farranda
anzas	abondamento	desbrumar	redomizar	regalable
ruíña	encrudo	marullecer	vesto	mallucar
esmerilidade	fraxibilidade	conversatorio	esvelar	lítico
desabastar	calugar	tintaxe	desembanar	cerraxeiro
éldro	amboraxe	labeiro	emborqueira	pelicento
rileiro	desengazar	reverderar	<u>pileiro</u>	ateimado

Test	Número de orde que ocupan as 20 pseudopalabras de cada subtest
1	3 6 10 12 14 16 17 23 25 30 33 36 40 45 46 49 50 53 55 56
2	3 5 9 11 12 13 15 17 18 19 21 32 37 39 46 48 49 51 53 58
3	1 2 5 6 8 11 12 13 24 31 32 35 36 37 44 45 47 58 59 60
4	2 4 7 13 18 21 22 23 25 30 31 34 36 37 39 43 45 46 57 58
5	2 3 4 15 16 17 18 19 21 31 32 34 40 41 49 52 53 55 56 58
6	1 2 3 5 7 10 15 19 22 23 24 26 27 37 38 43 49 50 55 60
7	1 4 14 19 24 25 26 29 30 31 32 37 41 44 45 46 56 57 58 59
8	10 15 17 22 23 25 26 29 30 31 32 37 41 44 45 46 56 57 58 59
9	1 8 9 10 14 15 17 19 21 22 31 38 41 42 43 49 50 57 58 59
10	4 5 6 13 16 17 20 25 30 31 32 34 38 39 43 44 47 49 51 55

ANEXO III

Formato inicial do test: Batería 1 e Batería 2

(Debido a cambios na configuración das páxinas, o formato orixinal aparece modificado neste anexo: na 1.^a páxina non aparece a que Batería pertence o protocolo e os tres últimos subtests inclúense en dúas páxinas en vez de nunha)

Nome e apelidos _____ Data de nacemento _____

Colexio _____ Grupo e curso _____ Data de hoxe _____

Coa miña nai e meu pai falo en: galego ☐ castelán ☐ nas dúas ☐

O meu pai traballa de: _____

A miña nai traballa de: _____

TEST nº 1

1. fito	16. moarí	31. maduración	46. bracar
2. ponche	17. mallarada	32. transvasar	47. chirloir
3. esbarecer	18. pecuniario	33. esmoular	48. continuar
4. redondear	19. hipótese	34. colección	49. grechar
5. campío	20. taxonomía	35. incluso	50. solegar
6. cediña	21. prolixo	36. nárbaro	51. resistencia
7. afectuoso	22. lambeteiro	37. esmigallar	52. amaño
8. asturiano	23. embornar	38. bovino	53. larpento
9. torradora	24. mantido	39. magoar	54. humidade
10. eslomecer	25. rendibilizar	40. rastrelar	55. presuntar
11. orzar	26. mesón	41. pinta	56. colomego
12. raciño	27. decasílabo	42. protozoos	57. solvente
13. achegar	28. planta	43. ambivalencia	58. virar
14. colzar	29. baixa	44. copa	59. temón
15. praza	30. croquela	45. radiosónico	60. baldreo

TEST nº 2

1. treboada	16. gabián	31. occidental	46. impermanente
2. anís	17. trepento	32. alfaiote	47. bisílabo
3. carcidume	18. teigume	33. onubense	48. cánodo
4. torrefacto	19. involvente	34. estelar	49. esloca
5. desfanar	20. abrasar	35. grupo	50. perspicacia
6. ligar	21. acelar	36. pasacalle	51. aramalla
7. portavultos	22. vexetar	37. cebrería	52. prosopopea
8. ética	23. orixe	38. foular	53. receiptamento
9. pasabrisas	24. carie	39. deslistar	54. figueira
10. agarimo	25. sobrecama	40. xeosinclinal	55. pero
11. trolo	26. umbilical	41. rousar	56. portavoz
12. arrancadoiro	27. impune	42. condición	57. heraldo
13. zaspán	28. afonía	43. infortunado	58. seritivo
14. lanar	29. pulseira	44. gástrico	59. montañés
15. refulgir	30. repartir	45. salaia	60. incorrer

TEST nº 3

1. desbartar	16. endivia	31. baldo	46. xeneroso
2. astroide	17. resguardo	32. cóncwego	47. desbrazamento
3. apisoar	18. serenidade	33. café	48. folla
4. perpiaño	19. intransferible	34. mediación	49. sacrificio
5. rabaliza	20. sabotador	35. pingoreteiro	50. vimbieira
6. anollar	21. fiador	36. esbullar	51. vítreo
7. diletante	22. demora	37. funxido	52. catalogar
8. rufalco	23. compacto	38. alustro	53. escuro
9. xabaril	24. tragalinguas	39. illar	54. asasino
10. chamuscar	25. publicar	40. poliedro	55. descomposición
11. desimpactante	26. conservación	41. rifar	56. nadar
12. postada	27. equivocación	42. condicionamento	57. momificar
13. carraxeiro	28. tunisiano	43. mariñao	58. áuleo
14. supernova	29. bater	44. embanelar	59. xordinte
15. estadillo	30. feitizo	45. aleísmo	60. maricidio

TEST nº 4

1. chavella	16. desorbitado	31. quecemanchas	46. recolmar
2. atego	17. agulla	32. felicitar	47. discurso
3. navío	18. felinio	33. rupestre	48. estrado
4. tubullo	19. puír	34. ornicultura	49. chula
5. insólito	20. alombar	35. avaro	50. pericia
6. montepío	21. bocellar	36. avanceiro	51. parolar
7. ouleiro	22. chalano	37. guilleiro	52. locomotriz
8. cancela	23. calenán	38. mafia	53. tuno
9. ortigueira	24. para	39. desbanar	54. urólogo
10. comigo	25. desdexar	40. contado	55. atopar
11. saurios	26. tradear	41. amañador	56. gateira
12. furtivo	27. inapetencia	42. regalicia	57. aromalía
13. fanego	28. araña	43. zaparallada	58. samergo
14. obediente	29. piña	44. crequenas	59. cotomelo
15. meigo	30. roubado	45. labiza	60. eufemismo

TEST nº 5

1. política	16. investidiario	31. seneridade	46. fedorento
2. helanismo	17. sárrico	32. merica	47. autorizar
3. moileiro	18. soneiro	33. bacía	48. mergullón
4. impudor	19. liañar	34. incepepe	49. anerxia
5. fusto	20. renovación	35. esbourar	50. empeñar
6. maseiro	21. indirixible	36. suspicaz	51. logaritmo
7. nuclear	22. lerchán	37. faltar	52. canceiro
8. esquilmo	23. acento	38. imperativo	53. cinleiro
9. incandescencia	24. galego	39. espanto	54. desenfrear
10. materialismo	25. bolsa	40. edulceración	55. labreiro
11. ungulado	26. chuvisca	41. consurcar	56. alcén
12. chadiano	27. desfacer	42. tremelucente	57. eclipsar
13. superpoñer	28. reticente	43. indíxena	58. ledicioso
14. asignación	29. cal	44. desprazamento	59. verdura
15. caletrán	30. navegar	45. hélice	60. escintilación

Nome e apelidos _____ Data de nacemento _____

Colexio _____ Grupo e curso _____ Data de hoxe _____

Coa miña nai e meu pai falo en: galego ☐ castelán ☐ nas dúas ☐

O meu pai traballa de: _____

A miña nai traballa de: _____

TEST nº 1

1. fervencia	16. arfar	31. primor	46. escolleito
2. papadeiro	17. verdello	32. puericultura	47. gardenia
3. esmerilidade	18. sensibilizar	33. proveedor	48. áureo
4. asentamento	19. rileiro	34. fastidiar	49. maristada
5. élidro	20. actividade	35. regaña	50. alivar
6. macico	21. corresco	36. apendicite	51. abondo
7. lunoso	22. franel	37. desabastar	52. desmedrar
8. compañía	23. presuntivo	38. impresentar	53. pírrico
9. dólar	24. abieiro	39. obseso	54. caracterizar
10. ruína	25. defecto	40. rebandar	55. anzas
11. golfo	26. amioeiro	41. posto	56. roca
12. padecer	27. bretoniano	42. premonición	57. cinseira
13. asento	28. megalómano	43. rexarcir	58. fronteira
14. relatividade	29. tremoia	44. xustificar	59. vítima
15. entear	30. interrogatorio	45. bando	60. esfingulación

TEST nº 2

1. engolar	16. hemiciclo	31. imbanación	46. ivar
2. constructivo	17. plural	32. zouneiro	47. contrabando
3. pita	18. plumaxe	33. gracia	48. proporcionar
4. alcasán	19. intervidura	34. refolear	49. arremedar
5. balear	20. sete	35. druida	50. convención
6. pomelo	21. evidenciar	36. revirado	51. insolvente
7. guanche	22. semáforo	37. reluscencia	52. sebo
8. promulgar	23. xene	38. terrazo	53. enlazar
9. mancomunar	24. nuril	39. fosfato	54. castañeta
10. abesouro	25. fraxibilidade	40. alargar	55. segmentar
11. raparigo	26. abondamento	41. repositar	56. piláculo
12. colérico	27. acelerar	42. detención	57. fiteira
13. cordialidade	28. perplexidade	43. indixencia	58. excluso
14. lagoar	29. desengazar	44. asebullón	59. amboraxe
15. filigrana	30. calugar	45. despicaz	60. cabestro

TEST nº 3

1. arrolar	16. servidume	31. cernidor	46. naturisto
2. dilucidar	17. perliscente	32. xermadólogo	47. mutuo
3. eminente	18. pesebre	33. avantaxar	48. deletrear
4. abeto	19. barítono	34. eivado	49. incultura
5. desvincular	20. intentar	35. hemípteros	50. debut
6. ponte	21. descentralizar	36. autosuficiente	51. cédula
7. sodomizar	22. hílico	37. chanella	52. concursar
8. convalecer	23. conversatorio	38. xullo	53. mencionar
9. nitrato	24. entrañas	39. corchea	54. vila
10. auniense	25. desbrumar	40. alteración	55. transformar
11. mecanizar	26. laniño	41. carrucén	56. esnourar
12. trepidante	27. bombardeo	42. desfigurar	57. proxulgar
13. axuntar	28. consenso	43. esculca	58. labeiro
14. seiscentos	29. reverderar	44. hemógrafo	59. marullecer
15. fanatidade	30. tintaxe	45. pardiño	60. propoñer

TEST nº 4

1. aranceira	16. inzo	31. publizar	46. salvavidas
2. delicado	17. raspadela	32. xurado	47. habilitar
3. destetar	18. dialectalismo	33. lei	48. félidos
4. incapacitar	19. apestadeira	34. volver	49. mounelo
5. refundir	20. caciquismo	35. curar	50. larugar
6. prospecto	21. tabanismo	36. estatuto	51. espeso
7. mutación	22. lardeiro	37. arrancada	52. falcatrueiro
8. loumado	23. emisor	38. redomizar	53. trasno
9. follieiro	24. sobre	39. tamén	54. sedición
10. barrentar	25. ancoraxe	40. desprovisto	55. amor
11. mandar	26. morriña	41. pileiro	56. distinción
12. nordeste	27. crocar	42. desembanar	57. quereza
13. estricnina	28. hedonismo	43. frisada	58. emborqueira
14. esvelar	29. socialdemocracia	44. respiro	59. vesto
15. tremulecente	30. chinchorro	45. extinción	60. aglomerar

TEST nº 5

1. regalía	16. cerraxeiro	31. súpula	46. escoupro
2. cromático	17. farranda	32. mallucar	47. artén
3. parricida	18. amarralla	33. chiscar	48. duriense
4. tendinismo	19. slogan	34. descéntrico	49. molella
5. demalar	20. lítido	35. damnificar	50. retablo
6. ateimado	21. camiseta	36. casca	51. devanzar
7. preso	22. predispoñer	37. velocista	52. miocardio
8. piar	23. miñense	38. fragulla	53. gravata
9. mil	24. vaixela	39. pelicento	54. carótide
10. firma	25. tullereiro	40. enterro	55. regalable
11. xardineiro	26. ciclón	41. discorrer	56. abandono
12. consecutivo	27. monllo	42. arras	57. crocodilo
13. caledo	28. beneficioso	43. desnilar	58. avalar
14. soterrar	29. investidura	44. prosbentos	59. zamburiña
15. tecnicismo	30. teimura	45. alternar	60. cátodo

Anexo IV

Táboa para calcula-la estimación do tamaño do vocabulario

Porcentaxe de palabras coñecidas a partir de Acertos e Erros

A= Palabras sinaladas correctamente; E= Pseudopalabras sinaladas

Entrar na columna da esquerda con A (número de acertos) e busca-la porcentaxe na columna correspondente ó valor E (erros)

A \ E	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
40	100	95	90	85	80	75	70	65	60	55	50	45	40	35	30	25	20	15	10	5
39	98	92	87	82	76	71	66	60	55	49	44	38	32	26	20	13	5			
38	95	89	84	78	73	67	61	55	50	44	37	31	24	17	10	1				
37	92	87	81	75	69	63	57	51	44	38	31	24	16	8						
36	90	84	78	72	65	59	52	46	39	32	24	17	8							
35	88	81	75	68	62	55	48	41	33	26	18	9								
34	85	78	72	65	58	51	43	36	28	20	11	2								
33	83	76	68	61	54	46	39	31	22	14	4									
32	80	73	65	58	50	42	34	25	17	7										
31	78	70	62	54	46	38	29	20	11	1										
30	75	67	59	51	42	33	24	15	5											
29	72	64	56	47	38	29	19	9												
28	70	61	52	43	34	24	14	4												
27	67	58	49	40	30	20	9													
26	65	55	46	36	25	15	4													
25	63	53	42	32	21	10														
24	60	50	39	28	17	5														
23	58	47	35	24	12															
22	55	44	32	20	7															
21	52	40	28	16	3															
20	50	37	24	11																
19	48	34	21	7																
18	45	31	17	2																
17	42	28	13																	
16	40	24	8																	
15	38	21	4																	
14	35	17																		
13	33	14																		
12	30	10																		
11	27	6																		
10	25	1																		
9	23																			
8	20																			
7	17																			
6	15																			
5	13																			
4	10																			
3	8																			
2	5																			
1	2																			

Anexo V

Instrucións iniciais para a aplicación do test

PRIMEIRAS NORMAS PARA A APLICACIÓN DAS PROBAS

Concédense 30 minutos para completa-los 5 subtests de cada Batería (uns 5 minutos para cada un dos 5 subtest).

Antes de empeza-la aplicación:

- 1) Explícaselles que se trata dunha proba para calcular cuántas palabras coñecen en galego.
- 2) Móstraselles a proba e indícaselles que non deben comezar ata que se lle diga.
- 3) Entréganselles as probas e indícaselles que deben completa-los datos que se lles piden ó principio, nun recadro.
- 4) Explícaselles no encerado cómo deben proceder, cun exemplo:
 - a) Deberán rodea-lo número daquelas palabras que estean seguros de coñecer.
 - b) Se non coñecen unha palabra, non deben rodea-lo número que lle corresponde:

Exemplo: “Se aparece a palabra “5. colexio”, seguramente todos a coñecemos. Por elo, rodeámo-lo número que leva diante.

Se aparece a palabra “relinga” e non sabémo-lo seu significado, ou non estamos seguros de coñecela, non rodeámo-lo número que leva diante.

No caso de que nos confundamos, riscámo-lo número cunha aspa. Deste modo, entenderase que esa palabra non a coñecedes. É coma se non rodearáde-lo número. No caso de que non coñezades unha palabra ou se pareza a unha que coñecedes, pero non é exactamente igual, non debedes sinalala e pasades á palabra seguinte”.

- c) Deben empezar pola palabra co número 1, seguir coa que ten o 2 e continuar así ata a que ten o número 15, logo o 16 ata a que ten o número 60, columna a columna.
- d) A continuación, deben pasar ó segundo subtest ata rematar co número 5.
- e) É importante recordarlles que se trata de sinalar soamente as palabras que están seguros de coñecer.
- f) Indícaselles que se lle vai a dar 30 minutos para completa-la Batería e que se lle indicará cada 5 minutos para axudarlles a axustarse ó tempo, que é o que aproximadamente lle ocupará cada subtest. No caso de que rematen antes de tempo, poden repasa-las contestacións.

NOTA: Aínda que se lle pasen as Baterías a todo o alumnado, ó recollelas, é preciso indicar no impreso que escolares presentan importantes dificultades na capacidade para ler e comprender o que len.

NORMAS PARA A APLICACIÓN DAS BATERÍAS

Concédense entre 25 e 30 minutos para completa-los 5 subtests de cada Bateria (uns 5 minutos por subtest).

Antes de empeza-la aplicación:

- 1) Explicarlles que se trata dunha proba para ver cuántas palabras coñecen en galego.
- 2) Mostrarlles a proba, indicándolles que non deben empezar ata que se lles diga:

AVouvos a repartir unha proba como esta (mostrándolles unha). Non a abrades ata que volo indique@.

- 3) Repártense os exemplares, colocándoos coa portada hacia abaixo, de forma que non se poda ve-lo contido. Ó mesmo tempo, asegurarse de que ninguén está sen lapis. A continuación decirlles:

APodedes darlle a volta ó caderniño, pero non escribades nada nel. Ben, primeiro poñémo-los datos que se vos piden na parte superior: nome e apelidos, data de nacemento, nome do colexio, curso e grupo, etc. (Ir ofrecéndolles indicacións ou aclaracións para completalos, se fixese falta)@.

- 4) Dáselle o tempo que precisen e seguimos. Explicarlles no encerado cun exemplo sobre cómo deben proceder:

AAntes de empezar, vamos a ver cómo tedes que facer para poder conseguir un bo resultado. Tedes que ler atentamente tódalas palabras de cada recadro. Comézase coa primeira palabra, se a coñecemos e estamos seguros de que é unha palabra en galego, sinalámola rodeando o número que ten diante; se non a coñecemos, non a marcamos e pasamos á seguinte. Vamos a facer algúns exemplos. (Para elo utilizamos os que se propoñen na proba).

Se aparece a palabra A5. Casa@, seguramente todos a coñecemos e estamos seguros de que é unha palabra en galego, aínda que tamén é unha palabra en castelán. ¿Rodeámo-lo número que leva diante? Si, porque a coñecemos.

Se aparece a palabra A 7. Vademécum@, como seguramente non a coñecemos ou se non estamos seguros de que sexa unha palabra en galego, ¿rodeámo-lo número que leva diante? Non.

Se aparece unha palabra como A16. Calle@, que non é unha palabra en galego, ¿rodeámola? Non.

Se aparece a palabra A22. Ralinga@, ¿que faremos?, ¿é unha palabra que coñecemos en galego? Ben, se non a coñecemos ou non estamos seguros, ¿rodeámola? Non.

Se aparece a palabra A42. Garneato@, ¿que faremos?, ¿é unha palabra que coñecemos en galego? Ben, se non a coñecemos ou non estamos seguros, ¿rodeámola? Non.

Se aparece a palabra A51. Encerado@, que é unha palabra tanto en galego coma en castelán, ¿rodeámola? Si.

*¿Está claro? Debedes ler cada palabra; se a coñecedes e vos parece que é unha palabra en galego, rodeáde-lo número que leva diante. Se a palabra se parece, pero non é a que coñecedes en galego ou non estades seguros, **non** rodeámo-lo número.*

No caso de que nos confundamos, riscámo-lo número cunha aspa, así (facelo no encerado) Con elo, entenderase que esa palabra non a coñecemos e que queda sen rodear. É coma se non rodearámo-lo número@.

É fundamental para unha correcta realización que lles queda claro isto, porque a tendencia que se detectou nas aplicacións piloto é a de rodear moitos ítems, coa esperanza de que así acertarán máis. Por esta razón, xeralmente produce o efecto desexado decirlles que as que rodeen mal, contan negativo e por iso é importante rodear só as que están seguros.

Seguidamente, continuamos explicando cómo deben proceder:

*A*Debedes empezar pola palabra co n1 1, seguir coa que ten o n1 2 e seguir así ata a n1 15, logo a 16 ata a que ten o n1 60, columna a columna. Cando rematedes, non pasedes ó seguinte test ata que volo diga. Vamos a esperar para rematar todos xuntos e empezar xuntos a seguinte. ¿Hai algunha dúbida? Se tedes algunha levantade a man e axudareivos (De non selo caso, dásele a sinal para que empecen).

É importante controla-lo desenvolvemento, ver que cumpren as instrucións, recordarlles que se trata de sinalar soamente as palabras que coñecemos, facendo algún comentario, se é o caso, do tipo de:

*A*Se vos sobra tempo, podedes repasa-las contestacións para aseguravos. Se alguén vai máis despacio, non quere decir que o faga peor@.

Con todo, se algún ou algunha vai demasiado lento, podemos animalo privadamente cun:

*A*Procura traballar un pouco á présa. Non é necesario que penses tanto cada palabra@.

Cando rematen todos a primeira subproba (recordar que se lle conceden uns 5 minutos, pero agárdase a que acaben todos), pídeselle que pasen á seguinte, decíndolles:

Agora vades a pasar á seguinte proba. Recordade que debedes empezar pola primeira columna coas palabras da 1 á 15, e logo coas outras tres columnas de palabras (sinalar). Lede cada palabra e se vos parece que a coñecedes e que é unha palabra en galego, rodeáde-lo número que leva diante. Se non estamos seguros, non rodeamos. Podedes empezar@.

Esta última instrución debemos repetila ó remate de cada proba, de forma que nos sirva para Aretelos@. Ter en conta que se detectou unha tendencia dos nenos e nenas, sobretodo dos máis pequenos, a contestar rápido e moitos ítems, co cal os resultados non son fiables.

Na aplicación da segunda Batería, deberemos recordarlle as instrucións.

NOTA: Aínda que se lle pasen as probas a todo o alumnado, ó recollelas, indicar no impreso cales presentan dificultades importantes na capacidade para ler e comprender.

Anexo VI

Instrucións definitivas para a aplicación do test T-VOGAL

NORMAS PARA A APLICACIÓN DAS BATERÍAS

Concédense entre 25 e 30 minutos para completa-los 5 subtests de cada Batería (uns 5 minutos por subtest).

Antes de empeza-la aplicación:

- 1) Asegurarse de que están o suficientemente separados, preparados e co material necesario (lapis ou bolígrafo).
- 2) Decirles que se trata dunha proba para ver cuántas palabras coñecen en galego.
- 3) Mostrarles a proba, indicándolles que non deben empezar ata que se lles diga:

A/Vouvos a repartir unha proba como esta (mostrándolles unha). Non lle dedes a volta á páxina ata que volo indique@.

- 4) Repártense os exemplares, colocándoos coa portada hacia abaixo, de forma que non se poda ve-lo contido. Ó mesmo tempo, asegurarse de que ninguén está sen lapis. A continuación decirles:

APodedes darlle a volta ó caderniño, pero non escribades nada nel ata que volo diga. Ben, primeiro poñémo-los datos que se vos piden na parte superior: nome e apelidos ..., data de nacemento ..., nome do colexio ..., curso e grupo ..., etc. (Ir ofrecéndolles indicacións ou aclaracións para completalos, se fíxese falta)@.

- 5) Dáselle o tempo que precisen e seguimos. Explicarlles no encerado cun exemplo sobre cómo deben proceder:

AAntes de empezar, vamos a ver cómo tedes que facer para poder conseguir un bo resultado. Tedes que ler atentamente tódalas palabras de cada recadro (sinalar). Comézase coa primeira palabra, se a coñecemos e estamos seguros de que é unha palabra en galego, sinalámola rodeando o número que ten diante; se non a coñecemos, non a marcamos e pasamos á seguinte. Vamos a facer algúns exemplos. (Para elo utilizamos os que se propoñen na proba).

Se aparece a palabra A5. Casa@, seguramente todos a coñecemos e estamos seguros de que é unha palabra en galego, aínda que tamén é unha palabra en castelán.) Rodeámo-lo número que leva diante? Si, porque a coñecemos (Para a Batería n1 2 podemos utilizar Avaca@).

Se aparece a palabra A 7. Vademécum@, como seguramente non a coñecemos ou se non estamos seguros de que sexa unha palabra en galego,) rodeámo-lo número que leva diante? Non (Para a Batería n1 2, utilizar Aroibén@, palabra que existe realmente: Acor vermella das nubes á saída e á posta do sol@).

Se aparece unha palabra como A16. Calle@, que non é unha palabra en galego, ¿rodeámola? Non (Utilizar Acarretera@ para a Batería n1 2).

Se aparece a palabra A22. Ralinga@ (é unha palabra inventada), ¿que faremos?, ¿é unha palabra que coñecemos en galego? Ben, se non a coñecemos ou non estamos seguros, ¿rodeámola? Non. (Para a Batería n1 2 podemos utilizar Atrebilla@, que tamén é unha palabra inventada).

Se aparece a palabra A42. Garneato@ (existe realmente: Aparte superior da gorxa@), ¿que faremos?, ¿é unha palabra que coñecemos en galego? Ben, se non a coñecemos ou non estamos seguros, ¿rodeámola? Non (Para a Batería n1 2 utilizar Avincallo@ que é unha palabra real: Aatadura de palla, xunco, etc.@).

Se aparece a palabra A45. Brola@, (é unha palabra inventada) ¿é unha palabra que coñecemos en galego? Ben, se non a coñecemos ou non estamos seguros, ¿rodeámola? Non. (Para a Batería n1 2 podemos utilizar Aremol@, que tamén é unha palabra inventada).

Se aparece a palabra A51. Encerado@, que é unha palabra tanto en galego coma en castelán, ¿rodeámola? Si (Para a Batería n1 2 utilizar Ames@).

Se aparece a palabra A60. Tremor@ (existe realmente: de tremer), ¿que faremos?, ¿é unha palabra que coñecemos en galego? Ben, se non a coñecemos ou non estamos seguros, ¿rodeámola? Non (Para a Batería n1 2 utilizar Arendo@ que é unha palabra real: de render).

*¿Está claro? Debedes ler cada palabra; se a coñecedes e vos parece que é unha palabra en galego, rodeáde-lo número que leva diante. Se a palabra se parece, pero non é a que coñecedes en galego ou non estades seguros, **non** rodeámo-lo número.*

No caso de que nos confundamos, riscámo-lo número cunha aspa, así (facelo no encerado) . Deste modo, entenderase que esa palabra non a coñecemos e que queda sen rodear. É coma se non rodearámo-lo número@.

É fundamental para unha correcta realización que lles queda claro esto, porque a tendencia que se detectou nas aplicacións piloto é a rodear moitos ítems, coa esperanza de que así acertarán máis. Por esta razón, é **necesario decirlles EXPLICITAMENTE** que as que rodeen mal, contan negativo e por eso é importante rodear só as que están seguros (esto xeralmente produce o efecto desexado).

Seguidamente, continuamos explicando cómo deben proceder:

Adebedes empezar pola palabra co n1 1, seguir coa que ten o n1 2 e seguir así ata a n1 15, logo a 16 ata a que ten o n1 60, columna a columna. Cando rematedes, non pasades ó seguinte test ata que volo diga. Vamos a esperar para rematar todos xuntos e empezar xuntos a seguinte. ¿Hai algunha dúbida? Se tedes algunha levantade a man e axudareivos (De non selo caso, dásele a sinal para que empecen).

É importante controla-lo desenvolvemento, ver que cumpren as instrucións, recordarlles que se trata de sinalar soamente as palabras que coñecemos, facendo algún comentario, se é o caso, do tipo de:

ASe vos sobra tempo, podedes repasa-las contestacións para asegurarvos. Se alguén vai máis despacio, non quere decir que o faga peor@.

Con todo, se algún ou algunha vai demasiado lento, podemos animalo privadamente cun:

AProcura traballar un pouco á présa. Non é necesario que penses tanto cada palabra@.

Cando rematen todos o primeiro subtest (recordar que se lle conceden uns 5 minutos, pero agárdase a que acaben todos), pídeselle que pasen ó seguinte, decíndolles:

AAgora vades a pasar á seguinte proba. Recordade que debedes empezar pola primeira columna coas palabras da 1 á 15, e logo coas outras tres columnas de palabras (sinalar). Lede cada palabra e si vos parece que a coñecedes e que é unha palabra en galego, rodeáde-lo número que leva diante. Se non estamos seguros, non rodeamos. Podedes empezar@.

Esta última instrución debemos repetila ó remate de cada proba, de forma que nos sirva para aretelos@. Ter en conta que se detectou unha tendencia dos nenos e nenas, sobretudo dos máis pequenos, a contestar rápido e moitos ítems, co cal os resultados non son fiables.

Na aplicación da segunda Batería, deberemos lerlle de novo as instrucións, aínda que o fagamos de forma menos pormenorizada.

NOTA: Aínda que se lle pasen as probas a todo o alumnado, ó recollelas, indicar no impreso cales presentan dificultades importantes na capacidade para ler e comprender.

Anexo VII

Formato definitivo do test T-VOGAL

(Debido a cambios na configuración das páxinas, o formato variou lixeiramente: o título de cada subtest aparece neste anexo máis xunto do listado de palabras que no orixinal)

T-VOGAL

BATERÍA nº 1

Nome e apelidos _____		Data de nacemento _____	
Colexio _____		Curso e grupo _____	
Data de hoxe _____			
Coa miña nai e meu pai falo en: galego <input type="checkbox"/> castelán <input type="checkbox"/> nas dúas <input type="checkbox"/>			
O traballo do meu pai é: _____			
O traballo da miña nai é: _____			

Exemplo:

5. casa	16. calle	42. garneato	51. encerado
7. vademécum	22. ralinga	45. brola	60. tremor

TEST nº 1

1. fito	16. moarí	31. maduración	46. bracar
2. ponche	17. mallarada	32. transvasar	47. chirlo
3. esbarecer	18. pecuniario	33. esmoular	48. continuar
4. redondear	19. hipótese	34. colección	49. grechar
5. campío	20. taxonomía	35. incluso	50. solegar
6. cediña	21. prolixo	36. nárbaro	51. resistencia
7. afectuoso	22. lambeteiro	37. esmigallar	52. amaño
8. asturiano	23. embornar	38. bovino	53. larpento
9. torradora	24. mantid	39. magoar	54. humidade
10. eslomecer	25. rendibilizar	40. rastrelar	55. presuntar
11. orzar	26. mesón	41. pinta	56. colomego
12. raciño	28. decasílabo	42. protozoos	57. solvente
13. achegar	28. planta	43. ambivalencia	58. virar
14. colzar	29. baixa	44. copa	59. temón
15. praza	30. croquela	45. radiosónico	60. baldreo

TEST nº 2

1. treboada	16. gabián	31. occidental	46. impermanente
2. anís	17. trepento	32. alfaiote	47. bisílabo
3. carcidume	18. teigume	33. onubense	48. cánodo
4. torrefacto	19. involvente	34. estelar	49. esloca
5. desfanar	20. abrasar	35. grupo	50. perspicacia
6. ligar	21. acelgar	36. pasacalle	51. aramalla
7. portavultos	22. vexetar	37. cebrería	52. prosopopea
8. ética	23. orixe	38. foular	53. receitaamento
9. pasabrisas	24. carie	39. deslistar	54. figueira
10. agarimo	25. sobrecama	40. xeosinclinal	55. pero
11. trolo	26. umbilical	41. rousar	56. portavoz
12. arrancadoiro	27. impune	42. condición	57. heraldo
13. zaspán	28. afonía	43. infortunado	58. seritivo
14. lanar	29. pulseira	44. gástrico	59. montañés
15. refulgar	30. repartir	45. salaíar	60. incorrer

TEST nº 3

1. desbartar	16. endivia	31. baldo	46. xeneroso
2. astroide	17. resgardo	32. cóncrego	47. desbrazamento
3. apisoar	18. serenidade	33. café	48. folla
4. perpiaño	19. intransferible	34. mediación	49. sacrificio
5. rabaliza	20. sabotador	35. pingoreteiro	50. vimbieira
6. anollar	21. fiador	36. esbullar	51. vítreo
7. diletante	22. demora	37. funxido	52. catalogar
8. rufalco	23. compacto	38. alustro	53. escuro
9. xabaril	24. tragalinguas	39. illar	54. asasino
10. chamuscar	25. publicar	40. poliedro	55. descomposición
11. desimpactante	26. conservación	41. rifar	56. nadar
12. postada	27. equivocación	42. condicionamento	57. momificar
13. carraxeiro	28. tunisiano	43. mariñao	58. áuleo
14. supernova	29. bater	44. embanelar	59. xordinte
15. estadullo	30. feitizo	45. aleísmo	60. maricidio

TEST nº 4

1. chavella	16. desorbitado	31. quecemanchas	46. recolmar
2. atego	17. agulla	32. felicitar	47. discurso
3. navío	18. felinio	33. rupestre	48. estrado
4. tubullo	19. puír	34. ornicultura	49. chula
5. insólito	20. alombar	35. avaro	50. pericia
6. montepío	21. bocellar	36. avanceiro	51. parolar
7. ouleiro	22. chalano	37. guilleiro	52. locomotriz
8. cancela	23. calenán	38. mafia	53. tuno
9. ortigueira	24. para	39. desbanar	54. urólogo
10. comigo	25. desdexar	40. contado	55. atopar
11. saurios	26. tradear	41. amañador	56. gateira
12. furtivo	27. inapetencia	42. regalicia	57. aromalía
13. fanego	28. araña	43. zaparallada	58. samergo
14. obediente	29. piña	44. crequenas	59. cotomelo
15. meigo	30. roubado	45. labiza	60. eufemismo

TEST nº 5

1. política	16. investidiario	31. seneridade	46. fedorento
2. helanismo	17. sárrico	32. merica	47. autorizar
3. moileiro	18. soneiro	33. bacía	48. mergullón
4. impudor	19. liañar	34. incepte	49. anerxia
5. fusto	20. renovación	35. esbourar	50. empeñar
6. maseiro	21. indirixible	36. suspicaz	51. logaritmo
7. nuclear	22. lerchán	37. faltar	52. canceiro
8. esquilmo	23. acento	38. imperativo	53. cinleiro
9. incandescencia	24. galego	39. espanto	54. desenfrear
10. materialismo	25. bolsa	40. edulceración	55. labreiro
11. ungulado	26. chuisca	41. consurcar	56. alcén
12. chadiano	27. desfac	42. tremelucete	57. eclipsar
13. superpoñer	28. reticen	43. indíxena	58. ledicioso
14. asignación	29. cal	44. desprazamento	59. verdura
15. caletrán	30. navegar	45. hélice	60. escintilación

T-VOGAL

BATERÍA nº 2

Nome e apelidos _____ Data de nacemento _____

Colexio _____ Curso e grupo _____ Data de hoxe _____

Coa miña nai e meu pai falo en: galego ☐ castelán ☐ nas dúas ☐

O traballo do meu pai é: _____

O traballo da miña nai é: _____

Exemplo:

5. vaca	16. trebilla	42. vincallo	51. encerado
7. roibén	22. carretera	45. remolo	60. rendo

TEST nº 1

1. fervencia	16. arfar	31. primor	46. escolleito
2. papadeiro	17. verdello	32. puericultura	47. gardenia
3. esmerilidade	18. sensibilizar	33. provedor	48. áureo
4. asentamento	19. rileo	34. fastidiar	49. maristada
5. élidro	20. actividade	35. regaña	50. alivar
6. macico	21. corresco	36. apendicite	51. abondo
7. lunoso	22. franel	37. desabastar	52. desmedrar
8. compañía	23. presuntivo	38. impresentar	53. pírrico
9. dólar	24. abieiro	39. obseso	54. caracterizar
10. ruíña	25. defecto	40. rebandar	55. anzas
11. golfo	26. amioeiro	41. posto	56. roca
12. padecer	27. bretoniano	42. premonición	57. cinseira
13. asento	28. megalómano	43. rexarcir	58. fronteira
14. relatividade	29. tremoia	44. xustificar	59. vítima
15. entear	30. interrogatorio	45. bando	60. esfingulación

TEST nº 2

1. engolar	16. hemiciclo	31. imbanación	46. ivar
2. constructivo	17. plural	32. zouneiro	47. contrabando
3. pita	18. plumaxe	33. gracia	48. proporcionar
4. alcasán	19. intervidura	34. refolear	49. arremedar
5. balear	20. sete	35. druida	50. convención
6. pomelo	21. evidenciar	36. revirado	51. insolvente
7. guanche	22. semáforo	37. reluscencia	52. sebo
8. promulgar	23. xene	38. terrazo	53. enlazar
9. mancomunar	24. nuril	39. fosfato	54. castañeta
10. abesouro	25. fraxibilidade	40. alargar	55. segmentar
11. raparigo	26. abondamento	41. repositar	56. piláculo
12. colérico	27. acelerar	42. detención	57. fiteira
13. cordialidade	28. perplexidade	43. indixencia	58. encrudo
14. lagoar	29. desengazar	44. asebullón	59. amboraxe
15. filigrana	30. calugar	45. despica	60. cabestro

TEST nº 3

1. arrolar	16. servidume	31. cernidor	46. naturisto
2. dilucidar	17. perliscente	32. xermadólogo	47. mutuo
3. eminente	18. pesebre	33. avantaxar	48. deletrear
4. abeto	19. barítono	34. eivado	49. incultura
5. desvincular	20. intentar	35. hemípteros	50. debut
6. ponte	21. descentralizar	36. autosuficiente	51. cédula
7. sodomizar	22. hílico	37. chanella	52. concursar
8. convalecer	23. conversatorio	38. xullo	53. mencionar
9. nitrato	24. entrañas	39. corchea	54. vila
10. auniense	25. desbrumar	40. alteración	55. transformar
11. mecanizar	26. laniño	41. carrucén	56. esnourar
12. trepidante	27. bombardeo	42. desfigurar	57. proxulgar
13. axuntar	28. consenso	43. esculca	58. labeiro
14. seiscentos	29. reverderar	44. hemógrafo	59. marullecer
15. fanatidade	30. tintaxe	45. pardizo	60. propoñer

TEST nº 4

1. aranceira	16. inzo	31. publizar	46. salvavidas
2. delicado	17. raspadela	32. xurado	47. habilitar
3. destetar	18. dialectalismo	33. lei	48. félicos
4. incapacitar	19. apestadeira	34. volver	49. mounelo
5. refundir	20. caciquismo	35. curar	50. larugar
6. prospecto	21. tabanismo	36. estatuto	51. espeso
7. mutación	22. lardeiro	37. arrancada	52. falcatrueiro
8. loumado	23. emisor	38. redomizar	53. trasno
9. folleiro	24. sobre	39. tamén	54. sedición
10. barrentar	25. ancoraxe	40. desprovisto	55. amor
11. mandar	26. morriña	41. pileiro	56. distinción
12. nordeste	27. crocar	42. desembanar	57. quereza
13. estricnina	28. hedonismo	43. frisada	58. emborqueira
14. esvelar	29. socialdemocracia	44. respiro	59. vesto
15. tremulecente	30. chinchorro	45. extinción	60. aglomerar

TEST nº 5

1. regalía	16. cerraxeiro	31. súpula	46. escoupro
2. cromático	17. farranda	32. mallucar	47. artén
3. parricida	18. amarralla	33. chiscar	48. duriense
4. tendinismo	19. slogan	34. descéntrico	49. molella
5. demalar	20. lítido	35. damnificar	50. retablo
6. ateimado	21. camiseta	36. casca	51. devanzar
7. preso	22. predispoñer	37. velocista	52. miocardio
8. piar	23. miñense	38. fragulla	53. gravata
9. mil	24. vaixela	39. pelicento	54. carótide
10. firma	25. tullereiro	40. enterro	55. regalable
11. xardineiro	26. ciclón	41. discorrer	56. abandono
12. consecutivo	27. monllo	42. arras	57. crocodilo
13. caledo	28. beneficioso	43. desnilar	58. avalar
14. soterrar	29. investidura	44. prosbentos	59. zamburiña
15. tecnicismo	30. teimura	45. alternar	60. cátodo

Anexo VIII

Táboas coas porcentaxes de sinalamentos dos ítems (subtests 1 e 3 da Batería 1 e subtests 3 e 5 da Batería 2)

Test 1.1.: Porcentaxe de elección das palabras na mostra Total (Batería 1, subtest 1)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	redondear	71	ponche	54	fito	20	pecuniario	2	esbarecer	55
2	asturiano	64	lambeteiro	78	campío	20	taxonomía	11	cediña	19
3	achegar	91	mantido	74	afectuoso	57	prolixo	10	eslomecer	10
4	praza	97	maduración	69	torradora	57	ambivalencia	2	raciño	11
5	mesón	80	transvasar	16	orzar	9	chirlomirlo	19	colzar	1
6	planta	65	incluso	51	hipótese	77	baldreo	25	moarí	1
7	baixa	96	esmigallar	57	decasílabo	46			mallarada	8
8	colección	86	bovino	50	magoar	65			embornar	9
9	copa	86	pinta	66	protozoos	36			rendibilizar	11
10	continuar	82	resistencia	76	solvente	15			croqueta	9
11	amaño	69	humidade	90	temón	57			esmoular	12
12	virar	85							nárbaro	1
13									rastrelar	27
14									radiosónico	21
15									bracicar	2
16									grechar	4
17									solegar	10
18									larpento	8
19									presuntar	13
20									colomego	4
Med.		80		62		42		11		12

Test 1.3.: Porcentaxe de elección das palabras na mostra Total (Batería 1, subtest 3)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	apisoar	66	chamuscarse	88	xabaril	73	perpiaño	3	desbargar	11
2	estadillo	5	resguardo	57	supernova	57	diletante	6	astroide	22
3	publicar	86	serenidade	81	endivia	33	sabotador	41	rabaliza	29
4	conservación	76	illar	51	intransferible	53	alustro	7	anollar	11
5	equivocación	75	condicionamento	34	fiador	48	mariño	4	rufalco	2
6	bater	81	xeneroso	95	demora	39	vítreo	5	desimpactante	22
7	café	96	sacrificio	80	compacto	56			postada	8
8	rifar	87	vimbeira	19	tunisiano	4			carraxeiro	17
9	folla	97	catalogar	57	feitizo	78			tragalinguas	44
10	escuro	91	decomposición	75	mediación	33			baldo	11
11	asasino	87	momificar	49	poliedro	37			cónchego	4
12	nadar	92							pingoreteiro	16
13									esbullar	13
14									funxido	9
15									embanelar	5
16									aleísmo	3
17									desbrazamento	14
18									áureo	1
19									xordinte	7
20									maricidio	7
Med.		78		62		47		11		13

Test 2.3.: Porcentaxe de elección das palabras na mostra Total (Batería 2, subtest 3)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	abeto	82	mecanizar	60	arrolar	63	dilucidar	2	auniense	2
2	ponte	96	servidume	63	desvincular	30	eminente	38	fanatidade	16
3	axuntar	37	entrañas	68	sodomizar	3	eivado	12	perliscente	3
4	seiscentos	96	avantaxar	53	convalecer	35	hemípteros	3	hílico	5
5	pesebre	79	autosuficiente	59	nitrito	39	esculca	4	conversatorio	34
6	intentar	78	dedfigurar	72	trepidante	48	cédula	11	desbrumar	14
7	bombardeo	86	incultura	41	barítono	38			laniño	5
8	xullo	88	debut	51	descentralizar	25			reverderar	4
9	alteración	72	mencionar	70	consenso	28			tintaxe	30
10	deletrear	79	transformar	84	corchea	84			cernidor	4
11	concurсар	85	propoñer	89	mutuo	56			xermadólogo	13
12	vila	92							chanella	7
13									carrucén	4
14									hemógrafo	16
15									pardizo	8
16									naturisto	12
17									esnourar	4
18									proxulgar	12
19									labeiro	5
20									marullecer	4
Med.		81		65		41		12		10

Test 2.5.: Porcentaxe de elección das palabras na mostra Total (Batería 2, subtest 5)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	preso	91	consecutivo	74	parricida	16	regalía	15	tendinismo	7
2	piar	69	soterrar	41	tecnicismo	18	cromático	22	demalar	6
3	mil	94	predispoñer	43	amarralla	7	escoupro	5	ateimado	33
4	firma	81	vaixela	63	slogan	40	duriense	5	caledo	6
5	xardineiro	76	ciclón	78	miñense	5	carótide	13	cerraxeiro	73
6	camiseta	93	velocista	40	monllo	13	cátodo	7	farranda	9
7	beneficioso	78	discorrer	63	investidura	50			lítico	8
8	chiscar	78	alternar	67	damnificar	26			tullereiro	4
9	casca	66	gravata	79	arras	33			teimura	25
10	enterro	81	crocodilo	85	miocardio	36			súpula	4
11	abandono	79	avalar	32	retablo	19			mallucar	5
12	zamburiña	77							descéntrico	19
13									fragulla	11
14									pelicento	4
15									desnilar	5
16									prosbentos	3
17									artén	7
18									molella	5
19									devanzar	12
20									regalable	45
Med.		80		60		27		11		15

**Test 1.1.: Porcentaxe de elección das palabras na mostra de 6.º de Educación Primaria
(Batería 1, subtest 1)**

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	redondear	71	ponche	56	fito	14	pecuniario	2	esbarecer	57
2	asturiano	62	lambeteiro	80	campío	14	taxonomía	8	cediña	21
3	achegar	92	mantido	70	afectuoso	56	prolixo	9	eslomecer	11
4	praza	97	maduración	70	torradora	58	ambivalencia	1	raciño	13
5	mesón	83	transvasar	12	orzar	7	chirlomirlo	15	colzar	1
6	planta	72	incluso	54	hipótese	69	baldreo	25	moarí	1
7	baixa	95	esmigallar	51	decasílabo	33			mallarada	8
8	colección	87	bovino	39	magoar	59			embornar	11
9	copa	87	pinta	66	protozoos	16			rendibilizar	10
10	continuar	83	resistencia	73	solvente	10			croqueta	10
11	amaño	74	humidade	88	temón	54			esmoular	11
12	virar	73							nárbaro	1
13									rastrelar	26
14									radiosónico	17
15									bracicar	3
16									grechar	4
17									solegar	7
18									larpento	9
19									presuntar	11
20									colomego	3
Med.		81		60		35		10		12

**Test 1.3.: Porcentaxe de elección das palabras na mostra de 6.º de Educación Primaria
(Batería 1, subtest 3)**

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	apisoar	71	chamuscar	91	xabaril	73	perpiaño	1	desbartar	14
2	estadillo	4	resgardo	52	supernova	59	diletante	5	astroide	25
3	publicar	86	serenidade	79	endivia	29	sabotador	46	rabaliza	32
4	conservación	77	illar	41	intransferible	47	alustro	7	anollar	7
5	equivocación	77	condicionamento	32	fiador	42	mariño	4	rufalco	1
6	bater	80	xeneroso	94	demora	28	vítreo	2	desimpactante	23
7	café	96	sacrificio	82	compacto	52			postada	10
8	rifar	87	vimbieira	14	tunisiano	3			carraxeiro	16
9	folla	98	catalogar	52	feitizo	77			tragalinguas	43
10	escuro	93	decomposición	76	mediación	33			baldo	10
11	asasino	84	momificar	41	poliedro	37			cónchego	5
12	nadar	92							pingoreteiro	14
13									esbullar	13
14									funxido	9
15									embanelar	4
16									aleísmo	2
17									desbrazamento	14
18									áureo	1
19									xordinte	3
20									maricidio	5
Med.		79		59		43		11		13

**Test 2.3.: Porcentaxe de elección das palabras na mostra de 6.º de Educación Primaria
(Batería 2, subtest 3)**

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	abeto	83	mecanizar	54	arrolar	61	dilucidar	1	auniense	2
2	ponte	96	servidume	57	desvincular	24	eminente	33	fanatidade	16
3	axuntar	33	entrañas	64	sodomizar	2	eivado	12	perliscente	3
4	seiscentos	98	avantaxar	45	convalecer	27	hemípteros	3	hílico	5
5	pesebre	79	autosuficiente	44	nitrito	25	esculca	3	conversatorio	37
6	intentar	79	dedfigurar	68	trepidante	40	cédula	11	desbrumar	9
7	bombardio	88	incultura	35	barítono	29			laniño	6
8	xullo	88	debut	45	descentralizar	20			reverderar	4
9	alteración	73	mencionar	69	consenso	18			tintaxe	27
10	deletrear	79	transformar	85	corchea	83			cernidor	3
11	concurar	86	propoñer	89	mutuo	54			xermadólogo	16
12	vila	94							chanella	6
13									carrucén	4
14									hemógrafo	15
15									pardizo	4
16									naturisto	13
17									esnourar	5
18									proxulgar	10
19									labeiro	7
20									marullecer	2
Med.		81		60		35		10		10

**Test 2.5.: Porcentaxe de elección das palabras na mostra de 6.º de Educación Primaria
(Batería 2, subtest 5)**

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	preso	93	consecutivo	70	parricida	10	regalía	15	tendinismo	6
2	piar	66	soterrar	16	tecnicismo	9	cromático	15	demalar	6
3	mil	96	predispoñer	37	amarralla	8	escoupro	6	ateimado	27
4	firma	86	vaixela	62	slogan	39	duriense	6	caledo	8
5	xardineiro	76	ciclón	74	miñense	5	carótide	12	cerraxeiro	69
6	camiseta	94	velocista	34	monllo	15	cátodo	6	farranda	9
7	beneficioso	76	discorrer	55	investidura	42			lítico	7
8	chiscar	76	alternar	63	damnificar	17			tullereiro	5
9	casca	72	gravata	75	arras	28			teimura	18
10	enterro	78	crocodilo	85	miocardio	25			súpula	5
11	abandono	78	avalар	25	retablo	33			mallucar	6
12	zamburiña	74							descéntrico	15
13									fragulla	12
14									pelicento	3
15									desnilar	7
16									prosbentos	4
17									artén	8
18									molella	8
19									devanzar	12
20									regalable	43
Med.		80		54		21		10		14

Test 1.1.: Porcentaxe de elección das palabras na mostra de 2.º de Educación Secundaria Obrigatoria (Batería 1, subtest 1)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	redondear	71	ponche	52	fito	27	pecuniario	2	esbarecer	53
2	asturiano	66	lambeteiro	76	campío	29	taxonomía	14	cediña	17
3	achegar	91	mantido	79	afectuoso	57	prolixo	11	eslomecer	9
4	praza	98	maduración	69	torradora	56	ambivalencia	3	raciño	9
5	mesón	77	transvasar	19	orzar	13	chirlomirlo	24	colzar	1
6	planta	58	incluso	46	hipótese	87	baldreo	25	moarí	1
7	baixa	96	esmigallar	63	decasílabo	63			mallarada	9
8	colección	86	bovino	63	magoar	72			embornar	7
9	copa	85	pinta	66	protozoos	60			rendibilizar	14
10	continuar	80	resistencia	80	solvente	20			croqueta	8
11	amaño	64	humidade	92	temón	61			esmoular	13
12	virar	79							nárbaro	1
13									rastrelar	28
14									radiosónico	26
15									bracicar	1
16									grechar	5
17									solegar	14
18									larpento	8
19									presuntar	16
20									colomego	5
Med.		79		64		49		13		12

Test 1.3.: Porcentaxe de elección das palabras na mostra de 2.º de Educación Secundaria Obrigatoria (Batería 1, subtest 3)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	apisoar	62	chamuscar	85	xabaril	73	perpiaño	6	desbartar	8
2	estadillo	5	resgardo	64	supernova	54	diletante	7	astroide	18
3	publicar	85	serenidade	84	endivia	39	sabotador	35	rabaliza	26
4	conservación	74	illar	63	intransferible	60	alustro	6	anollar	16
5	equivocación	72	condicionamento	37	fiador	56	mariño	3	rufalco	2
6	bater	83	xeneroso	96	demora	53	vítreo	9	desimpactante	22
7	café	96	sacrificio	78	compacto	60			postada	5
8	rifar	85	vimbieira	25	tunisiano	6			carraxeiro	17
9	folla	96	catalogar	63	feitizo	80			tragalinguas	45
10	escuro	69	decomposición	74	mediación	33			baldo	11
11	asasino	89	momificar	58	poliedro	48			cónchego	3
12	nadar	93							pingoreteiro	18
13									esbullar	12
14									funxido	9
15									embanelar	8
16									aleísmo	4
17									desbrazamento	14
18									áureo	2
19									xordinte	11
20									maricidio	8
Med.		78		62		51		11		13

Test 2.3.: Porcentaxe de elección das palabras na mostra de 2.º de Educación Secundaria Obrigatoria (Batería 2, subtest 3)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	abeto	80	mecanizar	67	arrolar	64	dilucidar	3	auniense	2
2	ponte	96	servidume	70	desvincular	37	eminente	44	fanatidade	19
3	axuntar	42	entrañas	72	sodomizar	4	eivado	12	perliscente	2
4	seiscentos	93	avantaxar	63	convalecer	44	hemípteros	4	hílico	5
5	pesebre	79	autosuficiente	77	nitrito	56	esculca	6	conversatorio	32
6	intentar	76	dedfigurar	77	trepidante	58	cédula	9	desbrumar	21
7	bombardio	84	incultura	49	barítono	48			laniño	4
8	xullo	89	debut	57	descentralizar	30			reverderar	5
9	alteración	71	mencionar	70	consenso	39			tintaxe	33
10	deletrear	78	transformar	83	corchea	87			cernidor	5
11	concurrir	83	propoñer	90	mutuo	59			xermadólogo	9
12	vila	91							chanella	7
13									carrucén	3
14									hemógrafo	17
15									pardizo	12
16									naturisto	11
17									esnourar	3
18									proxulgar	15
19									labeiro	3
20									marullecer	5
Med.		80		70		48		13		11

Test 2.5.: Porcentaxe de elección das palabras na mostra de 2.º de Educación Secundaria Obrigatoria (Batería 2, subtest 5)

N.º	Pal. nivel 1	%	Pal. nivel 2	%	Pal. nivel 3	%	Pal. nivel 4	%	Pseudopal.	%
1	preso	90	consecutivo	70	parricida	23	regalía	16	tendinismo	8
2	piar	69	soterrar	67	tecnicismo	29	cromático	27	demalar	5
3	mil	92	predispoñer	51	amarralla	7	escoupro	4	ateimado	40
4	firma	74	vaixela	64	slogan	41	duriense	4	caledo	5
5	xardineiro	77	ciclón	82	miñense	4	carótide	15	cerraxeiro	78
6	camiseta	91	velocista	47	monllo	11	cátodo	9	farranda	11
7	beneficioso	81	discorrer	72	investidura	58			lítico	11
8	chiscar	80	alternar	72	damnificar	37			tullereiro	3
9	casca	57	gravata	84	arras	38			teimura	32
10	enterro	83	crocodilo	85	miocardio	49			súpula	3
11	abandono	80	aval	41	retablo	46			mallucar	4
12	zamburiña	81							descéntrico	24
13									fragulla	10
14									pelicento	5
15									desnilar	4
16									prosbentos	3
17									artén	6
18									molella	3
19									devanzar	12
20									regalable	46
Med.		80		68		31		13		16

Anexo IX

Táboas de baremos en centís

BAREMOS para 6.º de Educación Primaria

Pc	Puntuacións en número de palabras estimadas			Pc
	BATERÍA 1	BATERÍA 2	TEST TOTAL	
99	10605	10826	10547	99
98	10046	10665	10224	98
97	9795	10414	9906	97
96	9692	10047	9581	96
95	9414	10001	9346	95
90	8311	9279	8589	90
85	7836	8455	7991	85
80	7526	8145	7661	80
75	7163	7784	7217	75
70	6599	7402	6959	70
65	6393	6908	6650	65
60	6186	6434	6392	60
55	5774	6187	6083	55
50	5361	5877	5774	50
45	5155	5568	5273	45
40	4640	5052	5000	40
35	4227	4640	4531	35
30	3815	4330	4134	30
25	3454	3712	3711	25
20	3093	3278	3351	20
15	2681	2681	2807	15
10	2062	2206	2340	10
5	1237	1587	1546	5
4	-----	1237	1368	4
3	1031	-----	1237	3
2	-----	883	957	2
1	618	618	647	1
N	313	313	313	N
Media	5342	5761	5551	Media
D.t.	2380	2596	2356	D.t.

BAREMOS para 2.º de Educación Secundaria Obrigatoria

Pc	Puntuacións en número de palabras estimadas			Pc
	BATERÍA 1	BATERÍA 2	TEST TOTAL	
99	12411	13247	12795	99
98	11391	12597	11670	98
97	11136	12336	11374	97
96	10665	12166	11342	96
95	10501	12047	11179	95
90	9764	11033	9934	90
85	9176	9749	9434	85
80	8661	9176	8970	80
75	8171	8867	8145	75
70	7836	8352	7635	70
65	7310	8042	7472	65
60	6908	7424	7155	60
55	6393	7011	6856	55
50	6083	6805	6599	50
45	5671	6495	6134	45
40	5361	6083	5774	40
35	4949	5470	5467	35
30	4722	5052	5047	30
25	4201	4536	4575	25
20	3774	4083	4021	20
15	3093	3655	3500	15
10	2475	2815	2799	10
5	2062	1974	2036	5
4	1856	1807	2036	4
3	1729	1650	1839	3
2	1264	1443	1445	2
1	878	748	786	1
N	262	262	262	N
Media	6174	6782	6478	Media
D.T.	2652	2953	2664	D.T.